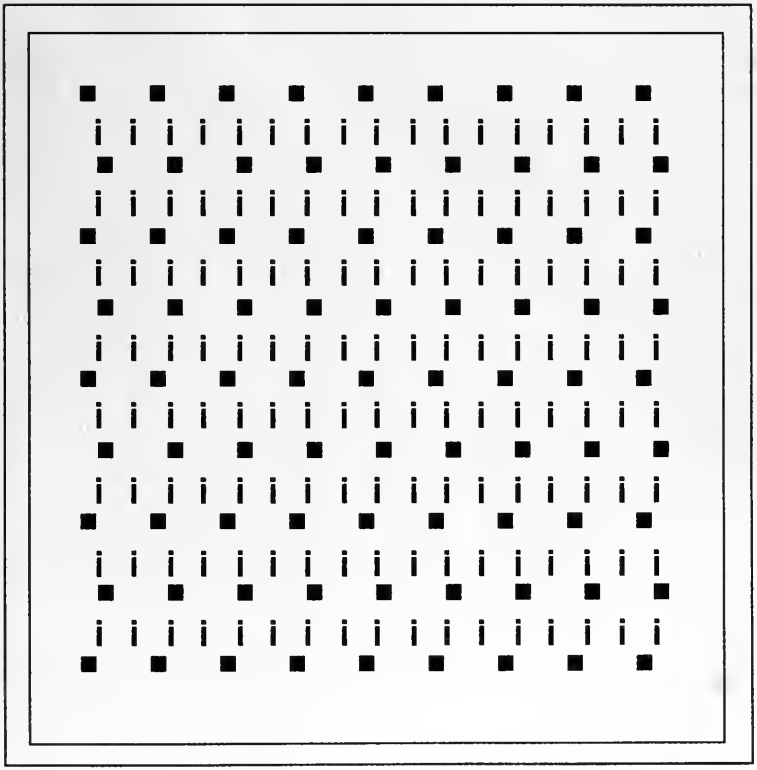
MAY 11 1989

IASSIST

QUARTERLY

VOLUME 12 FALL 1988 NUMBER 3



Printing compliments of The Rand Corporation

IASSIST QUARTERLY

VOLUME 12 NUMBER 3 Fall 1988

FEATURES

DEPARTMENTS

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

Digitized by the Internet Archive
in 2010 with funding from
University of North Carolina at Chapel Hill

http://www.archive.org/details/iassistquarterly123inte

Editorial Information

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST.**

Information for Authors

The **QUARTERLY** is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press **Manual of Style** or Kate L. Turabian's **Manual for Writers**. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2):77-82, March 1979.* If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor:

Walter Piovesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University
Burnaby, B.C., V5A 1S6 CANADA (01)604/291-4349 E-Mail: USERDLIB@SFU.BITNET

Book reviews should be submitted in duplicate to the Book Review Editor:

Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California
92713 USA. (01)714/856-4978 E-Mail: DTSANG@ORION.CF.UCI.EDU

Key Title: **Newsletter - International Association for Social Science Information Service and Technology**
ISSN - United States: 0739-1137 Copyright 1985 by **IASSIST**. All rights reserved.

Israeli Central Bureau of Statistcis

by Benjamin Lasman[1]
Directorate
Central Bureau of Statistics,
Jerusalem, ISRAEL

Introduction

The statistical service in Israel functions on a centralized basis and the Central Bureau of Statistics is the government agency responsible for collecting, processing and disseminating a variety of statistical data from surveys and administrative sources, and for carrying out national censuses. In the framework of the centralization of the statistical service, data collection for government needs is entrusted wholly to the Central Bureau of Statistics.

Methods of data collection, measurement techniques and procedures are the sole responsibility of this organization. The Statistical Law states that the Bureau shall be directed by scientific considerations only, and thus it is independent of any pressures on the methods it uses and procedures it employs in data collection, in the content of the information collected and the data released, or in the forms of data dissemination, etc.

The Bureau obtains information, the provision of which is mandatory, from suppliers of data such as households, firms, government and other public and private organizations. But the individual information obtained by the Bureau is kept confidential and no individual-level data can be given to any outside body whatever.

Population & Housing Census - CBS

Thank you for this opportunity to present, before this forum, the approach of the Central Bureau of Statistics in Israel to the theme of this Conference: "Public data: use it or loose it". I am sure that our approach is similar to that of other statistical offices, especially in countries with a centralized system of statistical services, but the specific experience may be different.

We at the Central Bureau of Statistics devote much thought and effort to the production of the large volume of data we collect, which is used extensively by government, public bodies, researchers, and other users. We are guided in this respect by the following principles:

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., May 26-29, 1988

1) The data collected by the Bureau should be the data which are needed by the users.

2) The data produced should be readily accessible to the users.

These two principles are the topics which I would like to discuss.

As to the first principle: in theory, public users can be asked to spell out explicitly their requirements for statistical information in advance of the time when they intend to use it. But in most cases, decision makers do not define in a proper manner the information they require, and in many cases they do not even know what data they will require at a future date. Our experience tells us that when a need for data arises, the data are required immediately, and as a statistical agency we also know that preparation of new statistics requires an extended period of time, sometimes a number of years.

Therefore the task of the Bureau is to use special means intended to determine if the Bureau is supplying the data which really are required, and if the figures published in over ten thousand pages annually are really needed. These means should also help to define the priorities to be put on the collection of various types of data. Hence, in order to ensure that the data collected are the data which are required, various tools have been developed and several bodies have been established for the purpose of being involved in the process of defining the data required and priorities.

One of the most important bodies, the functions of which are defined by law, is the Public Advisory Council for Statistics which brings together representatives of the main users of the information with the Central Bureau of Statistics, as the main producer of data. Among the representatives of the users are various government agencies, local authorities, trade unions, manufacturing associations, universities and research institutes, voluntary organisations and some independent experts. The Advisory Council has dealt with such subjects as development plans for various branches of statistics, censuses, classification – especially when a new subject becomes a policy issue, such as energy statistics and statistics on the environment.

A number of interdepartmental special advisory committees co-operate with the Bureau in addition to ad hoc sub-committees which the Advisory Council establishes. These interdepartmental special committees are set up to solicit comments and advice from users and researchers on specific statistical programs and projects. Among these committees are the Public Committee on the Consumer Price Index, on Input-Output Tables, on Labour Statistics, and special committees which advise on the planning and conduct of large surveys, such as Survey on Aging, Survey of University Graduates, Family Expenditure Survey, Survey of Travelling Habits and many others.

I would like to mention another aspect of the mutual contacts between the Bureau and the users. This is the function of "liaison officer" with the Central Bureau of Statistics, which has been established in various government offices and in other public organizations. These officials, as one of their functions, define the specific statistical needs of their organizations and submit them to the Bureau. They also know how to find out and use existing data. Individuals, or even special units, with such a function, exist in many departments and in some local authorities, and the Bureau encourages the development of these useful ties between public users and data producers.

After determining the types of data required, the relative priorities must be determined. In general, priorities are based on the importance of the data for planning and decision making as well as availability of funding possibilities. The influence of users on the order of priorities is

expressed mainly through the budget. An important part of the Bureau's budget, (close to 50%) is financed by government users, and this gives them a "say" on the subject of what statistical information will be collected, and some influence in determining priorities. In addition, the annual program of statistics is reviewed by a *"Steering Team"* composed of high government officials, who present their views on priorities and budget allocation to various projects.

I have reviewed here briefly the main ways in which users' requests and needs determine to a large extent the type of data collected by the Bureau. Nevertheless, it should be emphasized that the Central Bureau of Statistics does not wait for decision makers and planners to present their requests for specific statistical data, nor does it wait till financing has been assured. The Bureau tries to anticipate data needs and makes an effort to encourage users to define their requirements and to prepare tools that will be needed to obtain the data. This contributes to the existence of an orderly and comprehensive information system of statistical series. Thus, the Bureau tries to ensure that most data are available when they are required.

The second topic which I would like to discuss is the question how the Israeli Central Bureau of Statistics tries to make data readily accessible to users.

I do not intend to describe here the variety of conventional channels by which data are disseminated, of which printed publications still represent the most important medium. The Bureau's policy is to maximize availability and accessibility of data by developing new techniques, but this does not mean that one should disregard the importance of printed books, quick press releases, short bulletins, personal contacts with users by telephone and mail, etc. On the contrary, we have found that adopting modern technology in every stage of the collection and creation process improves many of the traditional forms of data dissemination. There are many examples of a substantial reduction in the time needed to bring data from the producer to the user by employing advanced computer techniques. One such example is the storage as microdata of a growing number of statistical series in data bases. While in the past, production of data in a format suitable for dissemination, such as camera-ready copies for printing, required weeks of planning and programming, today many statistical series are obtained instantly from data bases in the form of tables ready for distribution or publication. This is the case with statistics on national accounts, balance of payments, consumer price index, unit value indices of foreign trade, many series on labour, transport, industry, agriculture and others.

Electronic media of dissemination have a clear advantage over printed media, mainly in the speed and convenience with which users may access data. The Bureau has used this method in the past and uses it even more today by providing statistical information in machine readable form. One of the most important practices in this framework is the dissemination of anonymized microdata tapes to selected users, for research purposes. Great progress in this matter was made with the 1983 Population and Housing Census data. Within a very short time after the census was carried out, a series of public use tapes containing detailed census data were prepared and made available to users for processing. Two kinds of tapes were prepared for public use:

1. *Summary data tapes* based on the 100% enumeration, which include a limited number of variables and very wide geographic detail;

2. Two versions of *microdata tapes* which include practically the entire 20% enumeration sample.

The second category of tapes contain individual records of persons and households, with all

personal identification erased and other possibilites of identifying individuals eliminated. In spite of the relatively high cost of these tapes, users took advantage of the possibility to process census data themselves. The users of these data are mainly government agencies, municipalities, universities, research institutes and other large organizations. Due to the standard census tapes which the Bureau released for public use, availability of information from the last census is very widespread among the public users and the census serves as a source of very detailed information easily accessible to users for social planning and decision making and for research and teaching.

In this context I would like to mention the role of the Social Science Data Archive of the Hebrew University, which was the first organization to acquire the census tapes released for public use, in addition to many other tapes produced by the Bureau which contain data from most of the large surveys it conducts. The Social Science Archive plays an active role in further disseminating data produced by the Central Bureau of Statistics and in making these data more accessible than the format of the tapes released by the Bureau would otherwise allow. Special computer programs were developed by the Archive, including a data base which makes possible easy and immediate access to census data for each region, and enables users to obtain statistics on very detailed geographic bases, including the production of statistical maps. In this way the Archive serves researchers of the Hebrew University and of other universities and research institutes in Israel, as well as some other public users, by providing data for their use.

I would like to make two important comments on the dissemination of statistical data in the form of anonymized microdata: the first is, that we were very hesitant, some years ago, about distributing "microdata" and were apprehensive of the dangers and difficulties of processing data from these tapes outside the Bureau. These dangers included: deriving different figures from those published, using data for very small cells regardless of the problems of accuracy and statistical reliability that this may cause, potential breach of confidentiality, etc. But our experience has shown that by paying special attention to these problems in the planning and preparation of the microdata for public use, by providing proper and detailed documentation regarding the variables and categories included in the tapes, and by close contact with the users, mainly in the first stages of processing data, most of these dangers were avoided. Moreover, the usefulness of the data from various surveys and from the census was increased tremendously by the availability of the tapes.

The second comment which I would like to make is that the dissemination of microdata tapes by the Bureau for local processing requires highly skilled and well informed users, equipped with appropriate computer facilities. Therefore, the number of users of this kind of data is rather limited to large organizations which are interested and able to load their own computers with a great amount of data for long run and multi-purpose use. Hence this form of dissemination of machine readable data does not solve the need for a computerized system which would make it possible to expand the circle of users to the large number of new potential users of demographic, social and economic statistics, for business and administrative purposes whose demand for machine readable statistical information can not be satisfied by microdata tapes. This demand may be satisfied if the statistics were to be cheap, supplied in a simple format, and tailored for a wide range of uses. Unfortunately, in this respect, I am not in a position to present information from our experience, and I will be happy to listen to the achievements reported in the discussions during this Conference.

Thank you for your attention.

Public Data: Statistics Sweden

by Edmund Rapaport[1]
Statistics Sweden

As seen from the Swedish point of view, the title of this conference, "Public Data: Use It or Lose It", is as fateful and topical as it sounds. But this negativeness is not due to the lack of interested users – researchers and others. The problem is rather the apprehension in some quarters that there will no longer be the wealth of data available in the future as has previously been the case in Sweden.

Today, I would like to summarize the current situation in Sweden on the recording-keeping front, as concerns preserving material for research purposes, including statistical research. I assume that the general interest in the Swedish situation does not primarily stem from interest in Sweden as such, but rather from Sweden as an illustration, a potentially interesting example of record-keeping or archival problems in a small, modern nation, where "losing it" in the archival context was totally unheard of twenty years ago. The general operating rule has been that research material of potential value for research or in general for elucidating the future was to be saved – even the mere suggestion that such material should be destroyed was practically considered heresy in the recent past. The situation is different today, particularly with regard to materials stemming from population censuses, statistical materials collected for longitudinal studies or useful for such studies, and the like.

Although seldom explicitly expressed, the high value placed on archives has been deeply rooted in Swedish tradition, and the record-keeping system was – and still is – well organized and functional. In a recent report, a government commission summarized the general need for storing information carriers as follows: a) to satisfy the needs of citizens for information by ensuring public insight into the workings of society and thereby making it possible for individuals to participate in the democratic process; b) archives are also to serve as the memory of society and as a historical base for the furtherance of culture; c) archives are also a base in the search for information, both for advanced researchers and for others; d) archival institutions encourage effectiveness and rationalization by employing and developing methods for the optimal processing of information. Naturally, there are correlations and overlaps among these various objectives.

Negative arguments in counterpoint to these positive evaluations of the system of archives have increased in strength over the past twenty years, and had been aimed, if not at record-keeping as principle and as need, then at its scope and content. One troublesome factor is the cost factor. The amount of stored material is growing explosively, manual

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., May 26-29, 1988

processing is becoming relatively more expensive, and the costs of storage premises are rising substantially. Together with economic cut-backs in the public sector, these factors have given rise to demands for limits on record-keeping.

The popular expression "information society" is well suited to the present situation in modern Sweden. However, the extraordinarily fast growth in paper archives has made it necessary to destroy a great deal of this material. From 70–85% of all paper material in the national and municipal governments is destroyed, though usually on the condition that valuable information remains preserved in some other form, usually computerized. This policy is also followed by Statistics Sweden.

An argument practically unheard of twenty years ago was that the protection of privacy demanded the destruction of archival material. The background for this demand is the often rationally difficult to explain, yet still pronounced concern among the majority of the population about the consequences of storing information on individual persons. This concern has taken drastic expression on various occasions involving statistical studies, some of the most important of which are: the 1970 Swedish Census; a detailed study started in 1974 involving interviews on the living conditions of the population; a proposal for a register-based census in 1983; and, most recently in 1986, a longitudinal sociological study of a cohort of Stockholm school children, the so-called Project Metropolitan.

As concerns the preservation of archival materials produced by national and municipal governments, it has recently been "discovered" that this debate must also be put into the context of 200-year old Swedish principle of openness. This principle is regarded as an important pillar of Swedish democracy and is guaranteed in constitutional provisions. According to the principle of openness, anyone, including foreign citizens, may gain access to documents maintained by a public authority, without being required to give his or her name or purpose. Only those documents (or computerized records) that are explicitly exempt under the Official Secrecy Act are protected against this public observation. As long as all documents from public authorities were kept in storage, there was no reason to fear that the application of this principle would be obstructed because of missing documents. But now that the record-keeping requirement has been gradually dominated and openly questioned, at least as it concerns information on physical persons, there is the risk that the principle of openness may become at best a hollow entitlement.

So what is happening in Sweden today? The situation is unclear and there is rather intense analytical study and public debate underway.

Let us begin by stating that despite the extensive destruction of paper material mentioned above, the principle calling for the presentation of information on public activities in some form has remained undisturbed and will so continue until new guidelines are developed and passed into law. Within the national government, no documentation of value – unless it constitutes superfluous intermediary products in an archival-technical sense – may be destroyed without the permission of the responsible body, the National Swedish Archives and Record Office. 'Of value' is this context includes everything that may prove to be of value in the future. The National Record Office works to ensure that the time-honored Swedish principle of record-keeping is not relinquished. The only exceptions to the main rule of the sovereignty of the National Record Office are found in the Data Act, in force since 1974. This Act stipulates that the decision to destroy computerized registers of personal data rests with the Data Inspection Board, established under this law. The Data Inspection Board, however, is required to request an opinion from

the National Record Office before deciding to destroy material; in any case, the Board has not yet made any decisions conflicting with the interests of the National Record Office.

The current Swedish debate is characterized by two strong ideological or ethical streams: the strong popular interest in historical research, captured in the slogan "dig where you are standing" for various types of amateur research into "roots", on the one hand, and privacy interests that I would designate as a "noli me tangere" attitude, on the other. These currents flow in opposite directions. The third factor involved in public activities – that of economic cutbacks – constitutes the undercurrent.

The risks involved in the disappearance of valuable archival material, from the viewpoint of social researchers, are receiving increasing attention. The clearest example of this is found in public statements and accounts aimed at preventing the destruction of material seen as valuable for longitudinal medical, sociological, economic, and demographic research. Researchers are also worried about the increasingly acclaimed and even increasingly applied method of de-identifying information on individuals in order to protect their privacy. At the same time, it is apparent that for cost reasons, it is necessary to store material via computers – the method of record-keeping, however, that evokes the greatest apprehension from the standpoint of privacy. Manual processing, on the other hand, is becoming prohibitively expensive to repeat, if the paper forms (material) are preserved but not the computerized files.

In a very recent report submitted to the government on issues of record-keeping, demand is made for undiminished storage of government and municipal materials and for the expansion of record-keeping on the activities of businesses and other private institutions. At the same time, administrative simplifications are suggested, as are improved training of archival personnel and various measures for increasing effectiveness. Also needed are additional appropriations for archival activities. The proposal will receive thorough consideration.

In addition, a review of the constitutional rules on the principle of openness is underway, as is a review of the role of record-keeping for this principle; the findings concerning the latter will be published this autumn. A review of the Swedish Data Act has also been announced, in which demands for more restrictive record-keeping of personal information will undoubtedly be raised. Questions about personal privacy are politically sensitive, and there is no doubt that the Swedish government over the past 15 years has shown considerable sensitivity to demands for restrictions on the storage of personal information. This restrictive approach has not only been expressed in the terms of reference of various investigatory commissions, but also in concrete decisions.

Yet another commission has recently been appointed by the government to investigate ethical questions and rules of conduct in relation to the collection, use, and storage of individual information for social research. This commission will weigh the issues of destruction versus storage as well. It will examine an earlier proposal which raised questions about for long-term storage according to sampling principles; geographical and/or others.

Considering general policy development, I could also discuss the work of a number of other bodies involved in record-keeping issues, including a newly appointed commission to review the statutory regulation of national statistical activities. But I think that what I have already conveyed is sufficient to support the conclusion that we now find ourselves in a transitional period in the field of record-keeping. It should be stressed that Statistics Sweden, responsible for about 80 – 85% of all official statistics, is not exempt from the provisions of the Swedish Data Act, and

that the general development of archival policy pertains also to statistics.

The general policy of Statistics Sweden has always been, and still is, to support research outside the agency by allowing access to its registers, provided sufficient protection of confidentiality interests is guaranteed, in order not to harm the privacy of subjects involved.

Among the different developments which aim to facilitate access to our registers for research purposes, I would like to mention a new computerized model that has been developed at Statistics Sweden. This model makes it possible for a user to gain direct accesss to some microdata bases. But the user is only allowed to extract macrodata, which moreover has previously been controlled to prevent disclosure. Another development now underway is to produce public use tapes with so-called synthetic data, based on real empirical data.

I would like to end with some personal reflections.

For researchers, for analytically-oriented statisticians, and for historians, it is entirely clear that the preservation of documentation in archives is an inevitable precondition for their work, and is not only of interest to these disciplines, but also for to society as a whole. Record-keeping activities are conducted from a timeless perspective. The timeless perspective demands both an interest in the past as well as an interest in the future.

But limiting oneself to praising record-keeping is hardly a fruitful act today. Consideration must also be given to economic realities, since competition for public resources is tough. And we must take people's concerns about possible violation of their privacy quite seriously.

Solutions to the conflict are difficult to grasp, and the route development should take difficult to predict. It is easier to point out some features of the development and some desirable steps to be taken. For my part, I would like to summarize what I believe to be desirable goals in this field:

- Intensified research and development in order to be able to make greater use of computer technology, new computerized media, and other technology – such as modern storage technology – in the sphere of archives. The objective should be to decrease the space needed and to achieve other cost reductions. The method of record-keeping on a sampling basis may also need a closer look.

- Questions of documentation of computerized material must be given more attention and the demands must be intensified. At least in my experience from the field of statistics there are often deficiencies at this point in the process. Without documentation, the value of the stored material may decrease, perhaps approaching zero.

- Greater technical protection of stored material should be required. By this, I mean both conventional protection against access and protection in the form of encryption of materials, for example. An interesting model of encryption, aimed especially at protecting materials chosen for longitudinal studies has been developed at Statistics Sweden and will now be tested.

- The legal rules concerning access to stored material may also need review, and the requirements made tougher. The ethical guidelines that are applied in the professions in this and other fields may need codification, something that is occuring within various occupational groups, including within the statistical profession.

  Improved technical, legal, and ethical protection for archives may calm the public's fears.

- Information about archives, including statistical archives and their role in research, culture, education, politics, and the general development of society, must be intensified. Here the cooperation of all interest groups who benefit from archives – researchers, opinion makers, politicians – should be elicited. Information activities are also needed, for example, to illustrate the significance of longitudinal studies for progress in the medical and social spheres.

I am unable to judge whether my account here and my viewpoints are provincial or more generally applicable. In light of this doubt, I call for more international cooperation and exchange of information. Thus, I greet the initiative of this conference – with its provocative and current title – with great satisfaction.¤

Presenting Spatial Data: the Statistical Map As a New Practice

by Michal Peleg[1]
Social Sciences Data Archive
The Hebrew University of Jerusalem

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., U.S.A. on May 26-29, 1988

Introduction

What are the demographic characteristics of Washington, D.C., and how do they differ from those of Baltimore? What is the migration balance in the downtown area? Where do out-migrants tend to move to and what are the characteristics of mobile households? Is there a correlation between voting patterns and income level and what is the spatial distribution of this relationship? These questions represent statistical data on spatial relationships that govern much of our daily lives and occupy a variety of research fields in the social sciences.

The use of maps in presenting statistics, and the advantages and disadvantages of this method in comparison with more traditional methods will be discussed. I will present and analyse a series of maps of Jerusalem as a case study, and suggest a role for mapping and GIS (geographic information systems) in the data center environment. The function of the computer in producing maps will be mentioned briefly since this subject, which deserves a separate presentation, has already been discussed.

Why Produce A Statistical Map?

The impact of an article, proposal or report is strongly affected by the way the data is presented. Long tables of frequencies and percentages tend to wear the reader out after a few pages, whereas numbers are more easily interpreted when given visual support such as color or pattern graphics. Since the bulk of the information being provided is in the main statistical information, we can chose to provide it by traditional means: tables of frequencies, means, rates; or, if we are more sophisticated, we may instead prepare charts: pie, bar, line or three-dimensional graphs. Statistical charts and

graphs are very effective in creating interest and in appealing to the attention of a reader or an audience. If one variable in the distribution is spatial -- county, town, district, census tract etc. -- one has the additional option of mapping as a display technique. The other presentation techniques mentioned above lack any reference to spatial distributions, to a hierarchy of geographic divisions and the spatial relations among them.

A statistical map is merely another type of graphical presentation that takes advantage of a spatial variable and uses it in an unsymbolic way. The basis of the idea lies in the capacity of the map to act as a remarkably concise summary, to convey a great deal of information, the description and implication of which could otherwise be explained only in many pages of text[2].

A map has the advantage of showing a clear picture of the spatial distribution of a phenomenon and its implications. The following questions can best be answered using statistical maps rather than tabular displays. Do neighboring geographical units tend to have similar attributes? Can we identify the demographic, economic or urban effect of one area upon its neighbors over the years? What neighborhood dominates the quarter's average value and where is it located?

Maps that present these kinds of data are based on the existence of well defined geographic units which have some national, urban or ethnic homogeneity of population. Such units may be countries, cities, census tracts, blocks, regions, etc. While physical maps present a two-dimensional picture of the ground, statistical maps show mainly quantitative relations among the spatial units. Data suitable for representation on such a map are rates of marriage, divorce, birth, mortality and crime, rates of professionals or unemployed in the labour force, etc.

The Procedure of Map Making

The procedure of preparing a statistical map involves several stages and operations. It requires certain types of input, a process of method selection, map design, selection of equipment (hardware and software) and usually, several trial and error runs until the product meets the user's or the cartographer's demands. This procedure was used to produce a few statistical maps of Jerusalem. The object of these maps is to show the spatial distribution of two population characteristics: internal migration balance and socio-economic levels. Tables 1-3 in the appendix display the numeric data.

Migration balance was calculated from the 'Records of changes' by subtracting the number of out-migrants from the number of in-migrants, including first settlement of immigrants. Table 1 contains data for 3 consecutive years and the totals for those years. As can be seen, the net migration balance in Jerusalem is positive, though very small. Detailed examination of the data indicates a rather high variance among neighbourhoods. While the old neighborhoods, both Jewish and Arabic, do not show much movement, the new suburbs absorb a considerable amount of new population. Where are these spatial units located, what areas do they cover? How do these facts influence the urban texture of the city, its transportation system, location of new industrial plants, etc.?

Table 2 presents a subset of a study on the characterization and classification of urban geographic units in Israel. This work, carried out by the Central Bureau of Statistics, has been based on a multivariate analysis of 16 demographic and socio-economic variables derived from the 1983 census. By means of factor and cluster analysis, the values of these variables have been combined into a single measure for each geographical unit, then

converted to a standard distribution. Several maps of Jerusalem have been prepared to show the above data, to present the distributions and their meaning and to search for a relationship among them. (See maps 1–4 in the appendix).

Choosing the Right Method

The method and type of map must be selected according to the nature of the data, the medium of presentation, the purpose of the map, time and tools available for its preparation, and the audience for whom the map is intended. For any given set of data there is no absolute standard or criterion for selecting a particular type of map. With few exceptions, any set of spatially related statistical data can be portrayed in more than one map.

Non-quantitative maps are probably the simplest in their design. However, aspects of quantity are still the core of statistical mapping. Dickinson[3] divides quantitative statistical maps into 3 main types:

- Those in which quantities occur at a series of points
- Those in which quantities are contained within given areas
- Those in which quantities occur along a series of lines

This paper concentrates on the second type in which data, usually demographic, economic or social, have a spatial distribution that calls for emphasis. That is, a quantitative characteristic distinguishes one area from another and the rates of this differentiation are to be presented. Areal distributions can be presented by several mapping techniques: dot and isoline maps, repeated quantitative symbols, statistical diagrams and other methods. However, one of the more common ways to portray rates, based on clearly delimited areal units, is the choropleth map.

Choropleth maps are more common than any other type, although there are difficulties of a general nature associated with their design and use. The main difficulty lies in representing a quantity that is related to a given boundary-line area. Since the various areas in a map generally are not of equal size, the visual presentation of the areas distorts the distribution which is the subject of the map. Such bias is typical of choropleth maps and there are several techniques with which to deal with it. Another problem arises from a lack of information as to how items are arranged within a boundary-line. The larger the areas are, the greater the bias may be, and vice versa. Selecting smaller and more homogeneous areas (such as block or census tract) can reduce spatial distortion but it also means a larger data-set, an expensive base map, and sometimes results in an over-crowded map.

This method assumes uniform distribution of the subject variable across an area. The main problem is that the varying sizes of area have a substantial effect on the visual perception of quantities and their relationships. This difficulty can be overcome by displaying quantities as density per land unit or as a ratio, percentage, or per capita figure. This explains why a choropleth map is not suitable for representing purely numeric quantities. It can however, by itself or mixed with another technique, represent more than one variable. In such a case, one variable might be displayed by colors or shades and the second variable by crosshatching, repeated patterns, etc. An attempt to present more than 2 variables on a single map may result in a garbled and unclear picture. Obviously, there is a distinct limit to the amount of information which can be incorporated into a single chart and still be appreciated.

The choropleth method was chosen for the Jerusalem series of maps both to present single

characteristics and for the bivariate map. The reason for this decision lies in the nature of the data, our users' preference and available software.

Input

Two types of input are essential for creating a statistical map: a data-set that matches statistics to geographic areas and a base map suitable both to the selected method and map type.

Drawing a machine readable base map is one of the first steps in the preparation of any map type. The base map should be designed as a simple, outline map with a minimum of detail so that it can be further used for the production of other statistical maps of the same area. It normally consists of an outline of the contour of the geographical entity which is the subject of the map (city, county, country), and the polygons that represent its division into sub-areas. Such a map may include a few titles, symbols and other background details to make reading easier. However, unnecessary information may confuse the reader, detract attention from the main distributions and obscure the message of the map. Features such as rivers, place names and boundaries should be kept to a minimum, although they may usefully act as a "geographical framework" for the main subject. Extensively used base maps are now commercially available from government mapping agencies as well as from software houses. When the procedure is computer-aided, the base map is a file of digitally coded land and political boundaries, generated either by digitization or by scanning an accurate large map. In such a map each polygon represents an area according to a certain scale. Since one base map is normally used for producing numerous statistical maps, the investment in preparing a good map pays off.

Before preparing the base map for choropleth mapping, the appropriate geographic units must be selected. They should be small enough to be homogeneous and large enough to be of significance. The level of available data is another factor in selecting the geographical division. Jerusalem is divided into 3 hierarchical levels of areas - quarters, sub-quarters and statistical areas (census tracts). As a map of the quarters would appear to be too heterogeneous, while a map of the 155 statistical areas would be over-crowded, we chose sub-quarters as the unit of analysis. The series of Jerusalem base maps was created by digitizing the statistical area boundaries. Then, two less detailed maps were generated from the original one, one for sub-quarters and the other for quarters.

Interval Selection

In order to show the value of a variable (rates, percentages, or other statistics) for each individual area, the total range of values should be grouped into categories which are distinquished by different colors or shading. Whether the variable is discrete or continuous, the selection of the dividing points among groups considerably affects the significance of the map. The number of groups and their dividing points obviously depends on the range, shape and variance of the distribution. Methods of group creation vary from 'simple' thru automatic quantile groupings of the standard deviation. Maps 1 and 2 use the same intervals selected by quantiles with a modification that takes zero immigration balance as a dividing point. This results in a series of 5 groups: 3 categories for negative migration and 2 for positive values. In map 5, the two variables were grouped into 5 and 6 categories respectively which result in 30 different groups, each with a separate color and graphical presentation. It sounds confusing, but since each variable has a meaningful graphic scale the reader will immediately perceive the patterns and their meanings.

Pattern and Color Selection

Now that the intervals have been established, a suitable range of shading must be selected. Whether cross hatching or color, the principle is to arrange the patterns such that the lightest color or shade corresponds to the smallest magnitude with increasing densities showing successively higher values. In monochrome map production, this means arranging the density of lines in such a way as to give an optical effect from light to dark with respect to color or pattern. Each color or cross hatch pattern indicates an interval in a series of rates or percentages. Generally speaking, it is suggested that one not use too many different colors on any one map but rather one or two bright colors with their slightly muted tones. The use of color enhances the options of choropleth maps, and bright color maps attract the eye more than any other method; it should however be carefully applied.

Maps 1 and 2 show different color selections for the same data set with the same intervals. Map 2 was designed using two different colors. Positive migration is symbolized by the color red and its tones and blue corresponds to the negative migration balance. In Map 1 the intervals are distinguished by gradual green tones. Which map better reveals the facts and delivers the message is not a clear choice. Maps 3 and 4 continue the monochrome shading technique, using six different tones. Map 5 was produced by two consecutive computer runs: the SE score was plotted first, then re-plotted to present the migration balance by graphical patterns.

Trial and Error Process

Drawing a statistical map is a trial and error process. It is apparent by now that the rendering of statistics into pictorial form is far from being a simple procedure. It involves fundamental decisions on methods, equipment, techniques and other parameters of map design and normally entails laborious planning and plotting. Only after producing several drafts and tests does the product satisfy the professional user.

Data Centers, GIS and Statistical Maps

A considerable portion of the collection of a social science data archive consists of geographically oriented data, either microdata or small area statistics from national, regional or local surveys. Furthermore, several data centers have already built and maintain geographic information systems (GIS) for the processing of special purpose data sets. A GIS is normally created to handle the unique characteristics of geographic distributions and it can act as a very sophisticated device to make the most of this kind of data. Such a system is typically based on database management system (DBMS) software, using a hierarchical or a relational model for its schema. A well designed GIS also is capable of performing the conversion and translation of various spatial divisions and boundaries of the same areas. At the same time, we see an increasing effort to improve the presentation of data, especially by R&D and marketing professionals in national and local agencies. The expansion of computer based graphics, mainly LOTUS-like products, stirs the appetite as well as the expectations of these researchers and their audiences. Once they have become acquainted with colored bar and pie charts, they wish to see clear statistical maps when geographical distributions are being displayed. Since statistical maps are a natural output of GIS, they should be produced by data archives as an expansion of data processing services. It is necessary to enhance the GIS to include a coordinate system that corresponds to the geographical units involved. Recently developed software packages can do this, either on a personal computer (PC) or, for large and complicated systems, on a powerful mainframe. However, the success of such an activity

depends largely on the co-operation of a professional cartographer willing to share in the enterprise. The project as a whole can enhance both the geographic orientation and the related services offered by the data center.□

Acknowledgments:

To Amnon Shebo who has prepared the maps for this presentation.

References

1. Bachi, R., Cluff, V. V., and S. Ever-Handani. Geo-statistics in urban ecology - application to Toronto religious sub-populations. -in press April 1988.

2. Cuff, D. J. and M. T. Mattson. Thematic maps: their design and production. New York: Methuen, 1982.

3. Dickinson, G. C. Statistical mapping and the presentation of statistics. London: E. Arnold, 1973.

4. Marble, D. F., Calkins, H. W., and D. J. Peuquet. Basic reading in geographical information systems. Williamsville, NY: SPAD Systems Ltd., 1984.

5. Schmid, C. F. Handbook of graphic presentation. New York, NY: Ronald Press Co., 1979.

Table 1: Migration balance by sub-quarters (1983, 1984, 1985)

Sub-quarter		year			Total
Code	Name	1983	1984	1985	1983-85
	TOTAL	1,557	1,563	2,165	5,285
11	Ramot Eshkol	304	-452	-488	- 636
12	Beit Israel	-593	-720	-485	-1,798
13	Musrara	-284	-209	- 36	- 529
14	Center	-382	-254	-153	- 789
15	Rehavia	- 93	-236	-124	- 453
16	Nahlaot	-925	-541	-515	-1,981
17	Geula	-817	-489	-810	-2,116
21	Romema	- 22	247	121	346
22	Givat Shaul	15	543	1,432	1,990
23	Beit Hakerem	- 70	-354	-392	- 816
24	Givat Ram	50	14	47	111
25	Bait Vegan	75	-241	- 42	- 208
31	Qiryat Hayovel (S)	-404	-315	-139	- 858
32	Qiryat Hayovel (N)	-430	-262	-318	-1,010
33	Ir Ganim	-347	-216	-214	- 777
41	Gonen	-441	-422	-521	-1,384
42	Rassco	- 61	-307	-267	- 635
51	German Colony	-273	-212	-347	- 832
52	Talbieh	-100	-114	-135	- 349
53	Geulim	- 82	-126	82	- 126
54	Talpiot	-247	119	55	- 73
61	Christian Quater	- 14	1	57	44
62	Armenian Quater	13	- 3	43	53
63	Jewish Quater	130	101	-136	95
64	Moslem Quater	-333	- 58	- 50	- 441
71	Shuafat	508	- 23	241	726
72	Ramot Alon	2,836	3,387	2,686	8,909
73	Neve Yaacov	778	136	501	1,415
74	Givat Shapira	- 11	-138	- 11	- 160
75	A-Tur	-215	- 96	-109	- 420
76	Sheikh Jarrah	-121	- 93	- 70	- 284
81	Silwan	158	179	118	455
82	Sur Baher	49	288	56	393
83	East Talpiot	794	759	440	1,993
84	Gilo	3,236	2,200	1,648	7,084

Source: Jerusalem Statistical Yearbook, 1986

Table 2: Migration balance 1985, by sub-quarters - rates per thousand

Sub-quarter		Population	Migration Balance	Migration Per Thousand
Code	Name			
TOTAL		457,700	2,165	4.73
11	Ramot Eshkol	14,900	- 488	- 32.75
12	Beit Israel	24,000	- 485	- 20.21
13	Musrara	2,100	- 36	- 17.14
14	Center	5,000	- 153	- 30.6
15	Rehavia	7,800	- 124	- 15.9
16	Nahlaot	8,200	- 515	- 62.81
17	Geula	21,500	- 810	- 37.67
21	Romema	14,200	121	8.52
22	Givat Shaul	9,400	1,432	152.34
23	Beit Hakerem	16,900	- 392	- 23.20
24	Givat Ram	2,800	47	16.79
25	Bait Vegan	17,900	- 42	- 2.35
31	Qiryat Hayovel (S)	9,800	- 139	- 14.18
32	Qiryat Hayovel (N)	11,700	- 318	- 27.18
33	Ir Ganim	9,800	- 214	- 21.84
41	Gonen	24,000	- 521	- 21.7
42	Rassco	13,500	- 267	- 19.78
51	German Colony	11,100	- 347	- 31.26
52	Talbieh	3,800	- 135	- 35.53
53	Geulim	9,700	82	8.45
54	Talpiot	10,900	55	5.05
61	Christian Quater	4,500	57	12.67
62	Armenian Quater	2,000	43	21.50
63	Jewish Quater	2,200	- 136	- 61.82
64	Moslem Quater	17,600	- 50	- 2.84
71	Shuafat	32,400	241	7.44
72	Ramot Alon	20,100	2,686	133.63
73	Neve Yaacov	14,800	501	33.85
74	Givat Shapira	9,300	- 11	- 1.18
75	A-Tur	20,800	- 109	- 5.24
76	Sheikh Jarrah	7,500	- 70	- 9.33
81	Silwan	24,000	118	49.17
82	Sur Baher	15,900	56	3.52
83	East Talpiot	11,800	440	37.29
84	Gilo	23,900	1,648	68.95

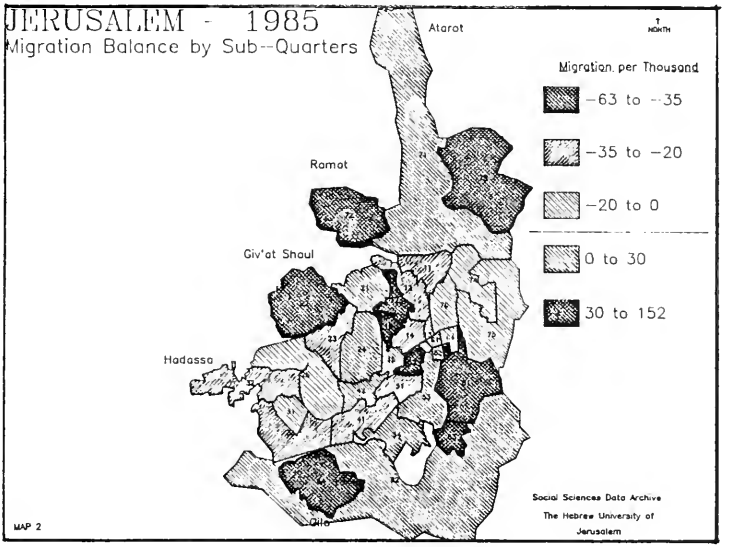
Source : Jerusalem Statistical Yearbook, 1986

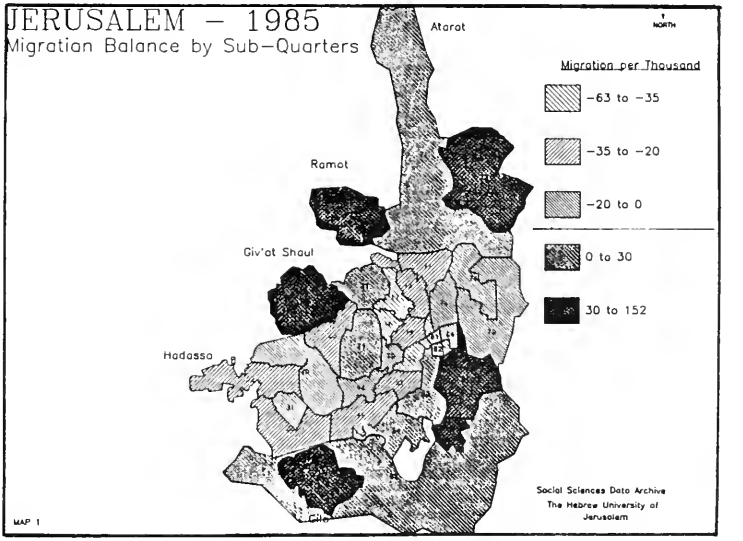
Table 3: Socio-economic scores of sub-quarters, 1983

Sub-quarter		Socio-Economic Score	Population
Code	Name		
11	Ramot Eshkol	0.71	14,803
12	Beit Israel	-0.18	23,168
13	Musrara	0.05	2,137
14	Center	0.25	2,367
15	Rehavia	0.84	7,490
16	Nahlaot	0.12	4,497
17	Geula	-0.10	21,417
21	Romema	0.01	12,341
22	Givat Shaul	0.17	6,030
23	Beit Hakerem	0.99	17,522
24	Givat Ram	0.67	1,584
25	Bait Vegan	0.51	17,248
31	Qiryat Hayovel (S)	0.51	9,018
32	Qiryat Hayovel (N)	0.42	11,924
33	Ir Ganim	0.26	9,442
41	Gonen	0.24	24,015
42	Rassco	1.19	13,466
51	German Colony	0.66	11,614
52	Talbieh	0.61	4,189
53	Geulim	0.37	8,945
54	Talpiot	0.52	9,037
61	Christian Quater	-1.01	4,322
62	Armenian Quater	-0.05	2,080
63	Jewish Quater	0.22	2,042
64	Moslem Quater	-2.66	17,102
71	Shuafat	-1.84	29,359
72	Ramot Alon	0.52	11,483
73	Neve Yaacov	0.45	13,111
74	Givat Shapira	1.03	6,876
75	A-Tur	-1.88	19,787
76	Sheikh Jarrah	-0.53	7,602
81	Silwan	-2.66	22,308
82	Sur Baher	-1.72	14,412
83	East Talpiot	0.69	9,419
84	Gilo	0.64	17,486

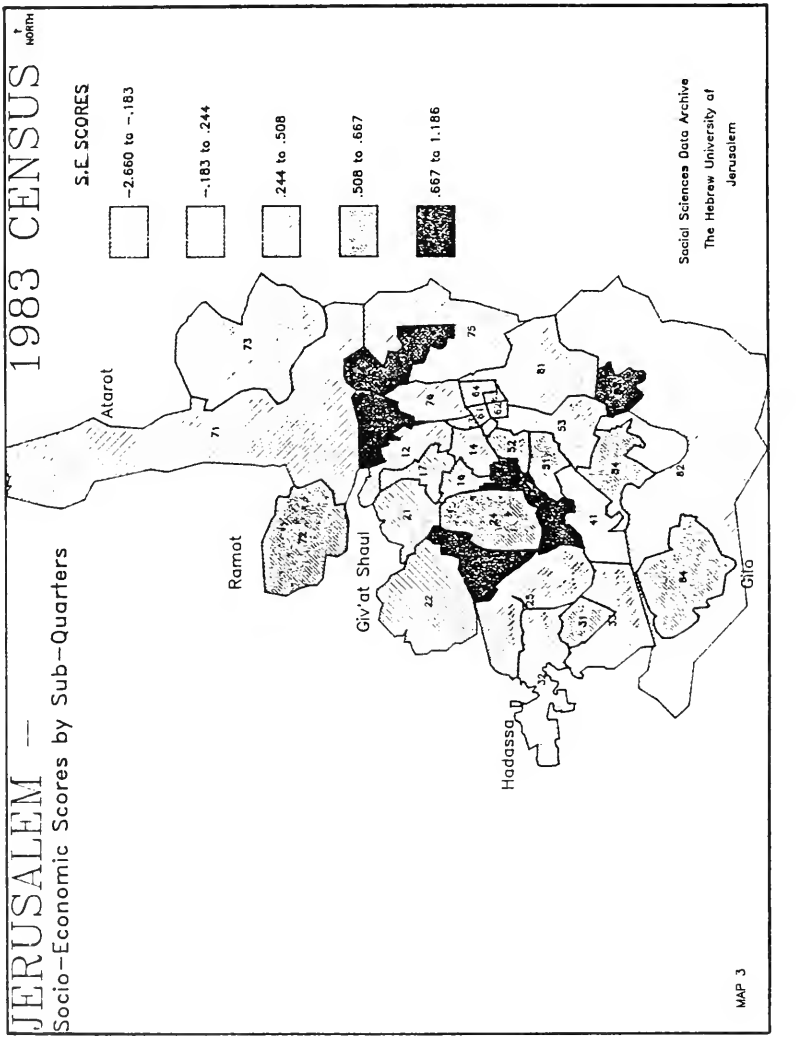
Source : The Central Bureau of Statistics

Maps 1 - 2: Migration Balance by Sub-Quarters, 1985

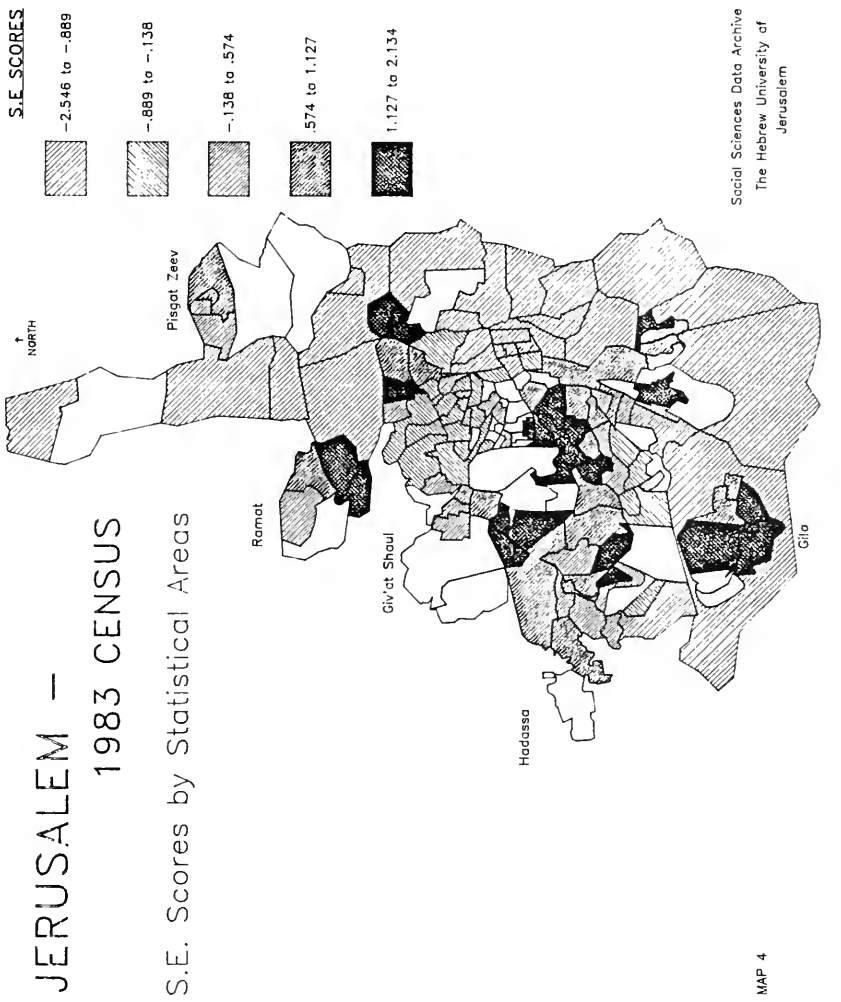




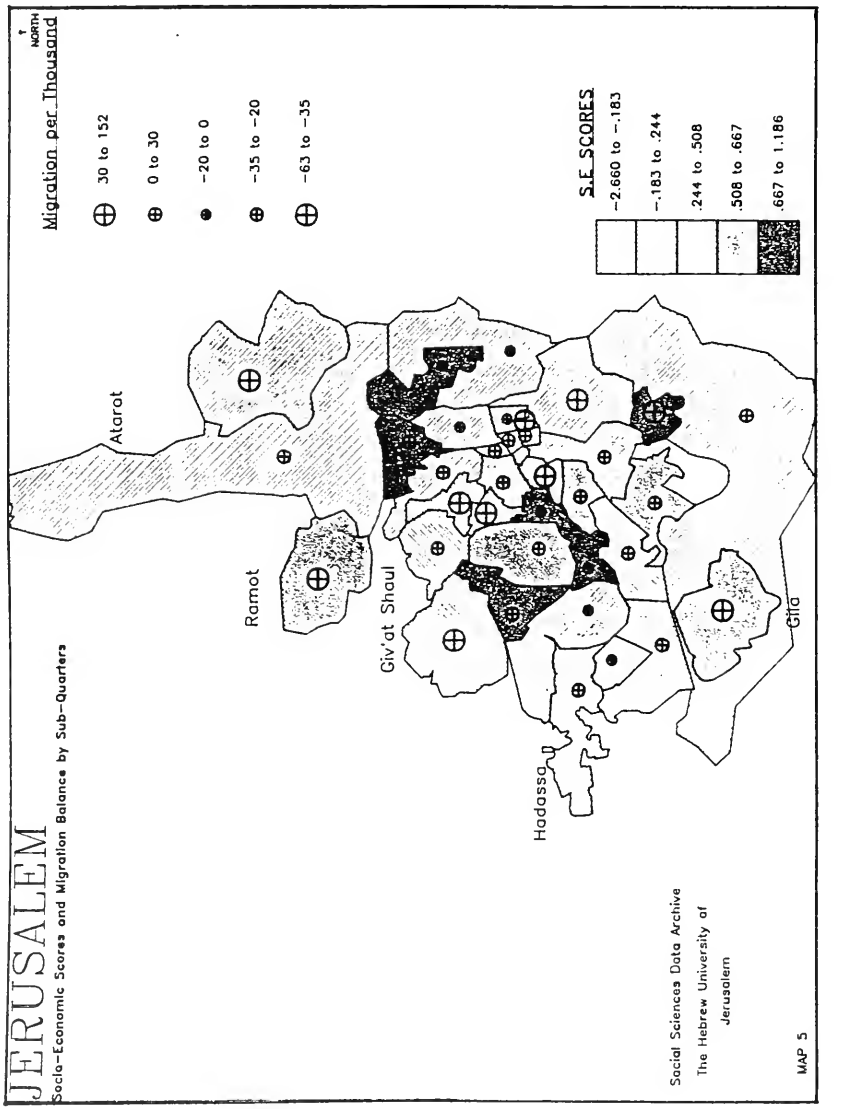
Map 3: Socio-Economic Scores by Sub-Quarters, 1983



Map 4: Socio-Economic Scores by Statistical Areas, 1983



Maps 5: Migration Balance and Socio-Economic Scores by Sub-Quarters



A Micro Based Tape Information System

by Martin Pawlocki
and Elizabeth Stephenson [1]
Institute for Social Science Research
University of California, Los Angeles

This paper is the third in a series describing the information systems of the ISSR Data Archive. The first two have dealt with an on-line bibliographic catalog of machine-readable data files and an in depth subject index, respectively (Stephenson & Bisom, 1986; Stephenson & Pawlocki, 1987). This paper deals with the management of administrative records of the contents of magnetic tapes using a database management system, dBase III. We will outline the decision-making process in the system design phase, the technical design of the system, data entry, system end products, and our future plans for augmentation of the system. We will present details of the system as it would be used by a small number of archive staff with a small to medium sized collection of materials. We will also outline the inter-relationship between this tape management system and other records and information maintained by the Data Archive.

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., May 26-29, 1988

Overview of archive information systems

The ISSR Data Archive was established in 1977, and since that time a variety of steps have been taken to make the retrieval and access of archive materials faster, easier, and more accurate. Our ability to achieve such a goal has improved dramatically with the acquisition of our IBM PC/AT. The software available and the relative inexpensiveness of storage and retrieval of information using the PC has made all the difference. The most important result has been that we are able to work within a tiered, or layered, environment and approach to information management. That is, we have been able to use the computing resources most appropriate to our own needs as well as those of the users. The Archive now maintains several types of on-line systems, accessible through different computing centers and facilities.

For the entire campus, the University Research Library (URL) maintains an on-line technical processing system called ORION, which can be used as an on-line library catalog of holdings. Within ORION, the Archive maintains its own database containing bibliographic details about studies in the Archive collection. This can be used as a browsing aid, or for locating specific titles.

This on-line catalog has some limitations for the description of machine-readable survey data.

The bibliographic entries cannot provide the abstracts necessary to describe in detail the content of files. Many surveys cover more topics than can be addressed by subject classification. Also, the on-line catalog is maintained on a mainframe computer, although subfiles may be downloaded to a personal computer-based database management system (DBMS).

To provide users with more detail about the content of individual data files, the Archive maintains subject oriented indexes which focus on specific areas (e.g., women's studies, ethnicity, health statistics), and contain abstracts and detailed indexing of the studies they include. Files are assigned up to 20 index terms appropriate to the subject area. The indexes are maintained for on-line searching within the Archive and will soon be available as part of a campus local area network, accessible to all users. Printed copies of the subject index are also available.

The library catalog and subject index are helpful only in identifying potentially useful data. Users of MRDF must examine variable lists and questionnaires to determine specific details about a data file. To address this need, researchers have access to the ICPSR's variable-based search facility operating under SPIRES on CDNet. In order to provide similar variable-level access to the contents of locally produced files, the Archive has developed a database containing the question text, variable names, and value codes for each question in a survey. At present, this database contains only the surveys conducted as part of the Los Angeles Metropolitan Area Surveys (LAMAS) or the Southern California Social Survey (SCSS).

When a file has been identified as being useful for research, a user needs technical information regarding the physical format and structure of the file, and its location on magnetic tape. For some time, the Archive maintained and provided this information on paper. With the acquisition of the microcomputer, and the availability of a DBMS, we began to explore the possibility of maintaining what was becoming a rather voluminous and unwieldy information system on computer. The following is a discussion of the planning, design and implementation of a tape information management system.

Preplanning and system design

Our first step, of course, was to discuss what we'd like in a system. This process took about three months of intensive effort, although we had discussed it previously as part of an ideal design for archive information management. It required that each staff member who was to use the system scrutinize every aspect of a variety of tasks, and the type of product or outcome to be produced. The staff using the system work in different areas of the archive, and with different sets of tasks and archival materials. We focused on many aspects of administrative records, reference needs, technical needs, records management for tapes, links to a campus information network, as well as the information needs of users.

The archivist viewed the system as potentially satisfying administrative and reference needs. There was a need for quick retrieval of technical and bibliographic information about every study held by the archive. The system would have to produce some paper products containing information about studies. Further, there was a need to be able to verify our holdings of particular studies by having a machine-readable shelf list.

Some of this information is traditionally stored in library catalogs. As previously mentioned, the archive does have a machine-readable

catalog using UCLA's on-line system, ORION. This catalog has proven itself useful for reference work, but the MARC-type record does not provide for the technical information that it is necessary to maintain about each file. This is especially true for multi-file studies, such as those which have have up to several hundred associated data sets. There is no mechanism in ORION with which to document all these files, and there appeared to be no way to combine bibliographic and technical information about several studies for users to access. We knew that a significant number of users knew which data file they wanted to use, and needed only the tape and file details in order to begin research projects. These needs would be better satisfied with a database management system.

The programmer's mission was to design a system that could be made accessible within UCLA's token ring network, SSCNet. This network uses Novell software (Advanced Netware) and an IBM Token Ring design. We wanted to make it possible for users to search the archive's tape system from anywhere on campus. We also wanted the system design to reflect the way in which we anticipated users would query the system, so that it could be self documenting. In addition, we wanted the programmer to design a system so that others familiar with dBase III might be able to interpret system bugs and understand the overall structure of the system. The programmer's design decisions are further outlined in the technical portion of this paper.

The technical assistant was to use the system from a task oriented viewpoint. Her needs focused on ease of entering, modifying, adding, and deleting information. As she would use the system repeatedly for the same procedures, it was important that she understand many details about the system and its limitations. The technical assistant would use the system to download the DCB information about each file of a study[2], and add bibliographic and other descriptive information, including subject terms, notes about file content, and file structure or format. She also needed the system for records management purposes. We wanted to be able to verify the contents of a single reel of tape, which might contain part of a large study, or many files of smaller studies. Other tape information needed for records management purposes included dates of file creation and tape cleaning, and the remaining space available on each tape. (In our facility, we follow a policy of filling all but 100 feet of a tape where possible. Knowing the number of unused feet of tape helps the technical assistant assign studies to archival storage on tape.)

For all of the Archive staff, there was a need for the system to be microcomputer based. Access to the mainframe computer (IBM 3090 with MVS and VM/CMS operating systems) is required for maintenance and copying of tapes, but it is an expensive medium for storage of administrative information. It is difficult to use the mainframe to produce the printed products we want, to link with other information systems maintained by the archive, nor is the mainframe accessible to all campus users. That is, not everyone on the campus has the financial resources to access the mainframe.

We also wanted to use the types of software that are available only in a microcomputer environment. While there are database management systems for mainframes, such as SPIRES, FOCUS, RAMIS, they are neither widely understood nor used by campus researchers. We wanted to avoid having to train users in understanding the structure and software of a DBMS. We felt that this would be more likely avoided in a microcomputer

[2] Editor's note DCB or data control block information is SAS nomenclature which includes record (or block) format (RECFM), physical record length (LRECL), and block size (BLKSIZE).

based situation. Microcomputers are also more attractive since the campus is being linked as a network via PC's, with the concommitant potential of making this system available to the largest number of users.

The following important information components were identified in the initial phase of system design: study title, principal investigator names, subject headings, study or accession number, tape volume id number, tape file numbers for each study, DCB information for each file, notes on individual file structure and format. We considered these to be the intellectual items we would need to produce a variety of system end products.

End products and system searching

The desired end products fell into three categories: paper, on-line internal, and on-line external (or system end products). The paper product is called a Tape Information Sheet. (An example is found in the Appendix). These have been part of the record keeping and information dissemination functions of the archive since it was established. The tape information sheet contains study title, names of principal investigators, major subject heading, tape volume id number, format of the tape (density, mode, parity, labeled/non-labeled), and the file numbers, dataset names and DCB information for each file associated with the study title. The purpose of the tape information sheet is threefold: 1) maintain a bibliographic shelflist; 2) maintain a tape number shelf list; 3) provide users with printed details about studies. A copy of the tape information sheet is maintained as part of a shelflist, filed by broad subject category for each study in the Archive collection, and also stored by tape volume id number so that we can verify the content of specific tapes.

We maintain files describing the content of each magnetic tape. As new studies are acquired, they are copied to Archive tapes which use a consecutive numbering system. A user accessible copy is stored at the computer center with a tape id number prefix "DTA" followed by a three digit number. As tapes are numbered consecutively, we can continue to use this system until we reach 999 reels of tape. An archival backup copy of the "DTA" tape is stored in a separate location and has the same consecutive numbering preceeded by the prefix "DTB". The tape information system stores information using the "DTA" tape number which is identical to the "DTB" tape number.

The third function of the tape information sheet is to provide users with a paper copy of all technical and bibliographic details required for the use of a data file. We found this to be essential in order that users have accurate information about the data they wish to use. Some computing centers have eliminated the need for this by cataloguing datasets, but our center is not set up in this way, and as we have stated earlier, the mainframe is not accessible to all. Further, the catalogued data sets cannot be easily linked to bibliographic and records management information, without significant systems programming.

Our other end products were largely focused on how the system would search for and retrieve information, and what information would be used in-house versus what would be publicly available. The actual process is described in the technical description below, but will be discussed in general terms here.

Basically, we wanted to be able to search the system in several ways: by title, principal investigator, subject, study number and tape number. We wanted to retrieve all information that met the search criteria. That is, we wanted

all editions of a study (for example, all versions of the American National Election Studies produced by ICPSR), or all years of a study with the same title (for example, Current Population Survey, March Annual Demographic File), but we did not want to have to specify, in the search, all editions or years, since we might not know that information. We also wanted to be able to view the complete list and select whichever files or studies were desired. In addition, we wanted to view a facsimile of the tape information sheet on the screen and, as described earlier, produce a print copy of the same information.

Some information and some types of searches were to be used only for internal purposes. For instance, several subject terms might be assigned to a study for use in searching, but only the broad intellectual category (e.g. mass political behavior) would appear on the printed tape information sheet. Also, users should not be able to perform searches by tape number, since they would be interested in specific files and would not need to know the content of a whole tape.

Data entry and authority lists

Once the system was designed and the programs written and debugged, we began an intensive period of data entry, beginning with the most heavily used types of files and entering all newly acquired data. Each tape is mapped or scanned using the mainframe, and this information is downloaded and reformatted. Bibliographic details are added through the use of screen templates, selection menus and prompts. The system was designed so that it would be easy to train ourselves and others in data entry, and when mistakes are made, to delete information or exit the system. The manner of downloading DCB information was a timesaving device and ensured that this information would be accurate and not subject to typing errors. The accuracy of the bibliographic information had been established using title/author and accession number authority lists. Subjects were assigned according to a pre-defined set of guidelines.

Subjects were meant to be broad and were to focus on categorization of the file rather than topical content. We also assigned geographic headings, acronyms of titles or principal investigators, and type of data descriptors such as "census". By and large, we tried to anticipate terminology that would be used by researchers when searching the system.

Future plans

This system will be very useful both for ourselves and for campus users until the collection becomes very large. At that time, the size of the database will be too large to accomodate the search pattern the system now employs. The campus is in the process of selecting a miniframe/microcomputer RDBMS. When this is acquired and put into place, we expect that we will be able to use its features to link all of the information systems we have developed, and perhaps provide linkage to the actual data from the files.

In the more immediate future, we would like to be able to permit users to download portions of their searches into personal files, and to produce a printed catalog of all our headings. We also expect to use the system with our detailed indexes and abstracts to create additional specialized indexes, but the design of such a

project has not been completed.

Technical descriptions

Until recently, much of the Archive administrative work was done on a mainframe computer. Over the last three years, we have been in the process of transferring this work to microcomputers. The project on which we are reporting started life as a way of producing tape information sheets on the microcomputer. When we noted the elements that would need to be entered, we realized that an opportunity to set up a microcomputer-based system to keep track of studies and tapes had presented itself. Once all the data are loaded, the system could become a catalog of our collection. So what was originally intended as an administrative system could also be used as a reference tool.

One of the first questions to be decided was which database management package to use. The options, because of availability, were dBASE III and R:BASE System V. dBase was chosen based on an analysis done by a library school student comparing the two packages. It was chosen because of its flexibility, power, speed, wide use on campus, and because the programmer was already very familiar with its programming language.

The primary concerns in developing the software were that it be user-friendly, menu driven, and most important, self-documenting. Once these issues were addressed in the design specifications, the system was fine-tuned in order to speed it up. This fine-tuning process is still going on.

The data base can be searched by title, subject, principal investigator, or study number. A record is retrieved if the search term is found anywhere in the field being searched. There is also an exact match search, and a left-adjusted search. That is, the search term must match starting at the leftmost character of the field being searched. For all types of searches, the results are displayed alphabetically by title. If the search is by principal investigator or subject, and more than one subject or principal investigator satisfies the search criteria, a list of results is displayed before individual titles are shown. The search can then be refined, or continued with the original search terms. The output from searching the database is the aforementioned tape information sheet.

The TAPES system is a relational database consisting of eight files, containing four types of fields: title, principal investigators, subjects, and the information relating to the computer files on tape (see the Appendix).

The title may contain up to 256 characters. Leading English articles are bypassed when the title sort key is created. The title sort key is a five digit code which, when accessed in ascending order, will keep the titles in alphabetical order. The codes were established with gaps to facilitate the insertion of new records. When the gaps are filled and the system slows down (a subjective judgement to be sure), a utility program is run to renumber the title keys. When stored in the file, titles are broken into lines of 71 characters to speed up the display of long titles. One of the principal investigator files contains the full text form of the principal investigator and a five digit code. The code is a consecutive accession number created when a new principal investigator (i.e. one not already in the file) is added. Each name appears only once in this file. Another principal investigator file contains the principal investigator code and the title sort key of the study to which this principal investigator is attached. As the system is presently configured, there can be up to nine principal investigators per study.

There is a similar arrangement for subjects. There are two files: one consisting of the full text form of the subject and a code, and the other containing the code and the title sort key associated with the study. As with principal investigator, up to nine subjects may be assigned per study.

Three files are used to keep track of the tape file information. The primary one contains data control block (DCB) information, the number of blocks and feet, (this information is downloaded from the mainframe), study number and a code for file format and file description (added manually). A second file contains information about the entire tape. Included are the number of files on the tape, the date the information was loaded into the system, the date of the last time the tape was cleaned, amount of space remaining on the tape, and whether more files can be stored on the tape. The third file consists of a note (if necessary) about the tape file. For example, the note field might contain the name of a state, or a description of the file.

When using the TAPES system as a means of accessing information about studies, the primary access point is the study number. The study number can be either the ICPSR number or a locally assigned number. For this later category, the study number can represent either an individual study or a group of studies. For example, instead of having each year of the March Current Population Survey (which we want to identify collectively) entered under its own individual ICPSR study number, a group number was assigned.

When used to get information about tapes, the primary access point is what is referred to in computer terminology as the tape volume serial number. Our tapes all have tape volumes starting with the letters "DTA" and numbered consecutively with a three digit number.

As we use the TAPES system, it is inevitable that problems will arise. The first modifications will be to fine-tune and correct any errors. The next major enhancement, however, will be to add abstracts to the studies.

Bibliography

dBase III; user manual. Culver City, CA: Ashton-Tate, 1984.

Guide to resources and services, 1987-88. Ann Arbor, MI: University of Michigan. Inter-university Consortium for Political and Social Research, 1987.

Liskin, Miriam. Advanced dBase III: programming and techniques. Berkeley, CA: Osborne McGraw-Hill, 1986.

Stephenson, Elizabeth A. and Diane Bisom. *Using ORION to CREATE an on-line catalog of data archive holdings.* Paper presented at the IASSIST Conference, Los Angeles, CA., May 21-25, 1986.

Stephenson, Elizabeth A. and Martin Pawlocki. *Index of machinereadable data files for women's studies.* Paper presented at the IASSIST Conference, Vancouver, B.C., May 19-22, 1987.

TITLE
TITLE SORT KEY
STUDY NUMBER
Principal Investigator (full)
Principal Investigator code
title sort key
Principal Investigator code
Principal Investigator (code)
Subject Heading (full)
subject heading code
subject heading code
title sort key
Subject Heading (code)
Tape Info
DTA number
DTA number
study number
File Info
DTA number / file number
DTA number / file number
File Description

2

***** THESE DATA ARE MADE AVAILABLE FOR USE EXCLUSIVELY *****
***** AT UCLA, AND MAY NOT BE REDISTRIBUTED *****

Miller, Warren E.

AMERICAN NATIONAL ELECTION STUDY, 1986

STUDY NO: I8678 VOL ID: DTA446, DTA499

TRACK: 9 DENSITY: 6250 LABEL: SL MODE: EBCDIC PARITY: ODD

FILE DESC.	FILE NO.	DSNAME	RECFM	LRECL	BLKSIZE	FILE FORMAT	APPROX NO. OF RECORDS
Dictionary (DTA446)	5	DI8678.CPS Post Election Survey, April 1987 (CPS version)	FB	80	1600	OSIRIS	870
Data (DTA446)	6	DA8678.CPS Post Election Survey, April 1987 (CPS version)	FB	1227	4908	OSIRIS	2174
Dictionary (DTA499)	7	ODICT.S8678 Post Election Survey, 1986 (ICPSR Version)	FB	80	1600	OSIRIS	810
Data (DTA499)	8	ODATA.S8678 Post Election Survey, 1986 (ICPSR Version)	FB	1220	29280	OSIRIS	2172
Codebook (DTA499)	9	OCDBK.S8678 Post Election Survey, 1986 (ICPSR Version)	FB	80	30000	OSIRIS	20062
Codebook (DTA499)	10	CB8678 Post Election Survey, 1986 (ICPSR Version)	FB	80	30000	LISTED	31312
Data Def. (DTA499)	11	SP8678.LRECL Post Election Survey, 1986 (ICPSR Version)	FB	80	30000	SPSS	3937
Data (DTA499)	12	FQ8678 Post Election Survey Frequencies (ICPSR Version)	FB	132	29964	LOGICAL	2156

3

```
                    WOULD YOU LIKE TO WORK WITH THE:

                      (S)tudy Records
                      (T)ape Records
                      (D)atabase Search
                      (M)aintenance
                      (E)xit
                      (R)eturn to DOS

                    Please enter the appropriate letter [S]

(A)dd,  (M)odify,  (D)elete,  (V)iew,  (P)rint,  (E)xit [A]

                          Study Number,  (E)xit [TEST    ]
```

4

eg.4

```
                    Study Number = TEST
***** IF YOU DON'T WANT TO ADD THE STUDY, LEAVE THE TITLE BLANK *****

Title  Test Record

P.I. 1 Test P.I. number 1
P.I. 2 Test P.I. number 2
P.I. 3
            ***** Don't worry, you can enter more later *****

Subject Term 1 Test subject number 1
Subject Term 2 Test subject number 2
Subject Term 3
            ***** Don't worry, you can enter more later *****
```

5

eg.5

```
STUDY = TEST

Tape Number, (E)xit  DTA001
File Number (either one [1] or a range [1 - 3]),  (E)xit  1-5
```

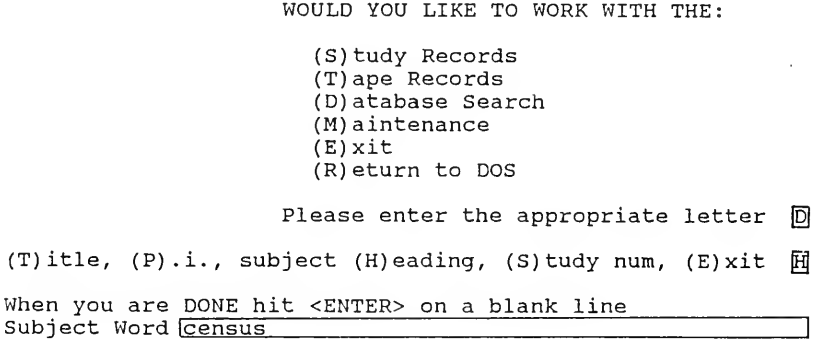
```
                STUDY: TEST    PART: [ ]      DTA001 FILE: 1

DSN = LIB.PUBLIC.A74
RECFM = FB
LRECL =   800
BLKSIZE =  2400

File Description: (D)ata, (C)odebook, d(I)ctionary, (J)cl, (B)lank, (E)xit [D]

If there is no file description note, leave the line blank
NOTE: [A test record note                                                  ]


File Format:
(O)siris, (S)pss, spss(X), s(A)s, (L)ogical, (C)ard, listed-to-(T)ape [L]
```


WOULD YOU LIKE TO WORK WITH THE:
(S)tudy Records
(T)ape Records
(D)atabase Search
(M)aintenance
(E)xit
(R)eturn to DOS
Please enter the appropriate letter D
(T)itle, (P).i., subject (H)eading, (S)tudy num, (E)xit H
When you are DONE hit <ENTER> on a blank line
Subject Word census
FOR MORE INFORMATION, TYPE ? AND HIT <ENTER>

THE FOLLOWING SUBJECT HEADINGS MATCH YOUR SEARCH TERM

Census
Census - 1790-1970
Census - 1900
Census - 1940
Census - 1950
Census - 1960
Census - 1970
Census - 1980
Census - 1990

Hit <ENTER> to CONTINUE, (*E)xit, or key in a NEW SUBJECT HEADING
Subject Heading census - 1980

SUBJECT HEADING = Census - 1980

STUDY ID		
I7789	Census of Population and Housing, 1980 (U.S.):	Census Software Package (CENSPAC) Release 3.1
I8471	Census of Population and Housing, 1980 (U.S.):	County Migration by Selected Characteristics, 1975-1980
I8405	Census of Population and Housing, 1980 (U.S.):	MARF 3
I8323	Census of Population and Housing, 1980 (U.S.):	MARF 5
I7854	Census of Population and Housing, 1980 (U.S.):	P.L. 94-171 Population Counts
I8101	Census of Population and Housing, 1980 (U.S.):	PUMS A Sample: 5% Sample
I8664	Census of Population and Housing, 1980 (U.S.):	PUMS American Indian Supplementary Questionnaire
I8170	Census of Population and Housing, 1980 (U.S.):	PUMS B Sample: 1% Sample
I7781	Census of Population and Housing, 1980 (U.S.):	Richmond Dress Rehearsal
M140	Census of Population and Housing, 1980 (U.S.):	Spanish Surname List
I7941	Census of Population and Housing, 1980 (U.S.):	STF 1A
I7975	Census of Population and Housing, 1980 (U.S.):	STF 1B

Key STUDY ID, hit <ENTER> to continue, or (E)xit [ ] THERE ARE MORE HITS

The Medieval and Early Modern Data Bank at Rutgers

by Rudolf M. Bell
and Martha Carlin [1]

Imagine that you are in Siena in the year 1368. You and your two brothers come from a wealthy family and have just been deposed from the city government by a revolt. After several days in hiding, you find that the new Sienese authorities will allow you to emerge and "remain at peace" in return for a fine of 100 gold florins. Now for the 64,000-florin question: how substantial was this fine? How many loaves of bread could you buy with it? Would it pay a craftsman's wages for a year? Or the rent of a substantial house? Could you have lived in luxury on it for a year, or in poverty for a month? If you chose not to pay, but rather to flee the city and await better times elsewhere, say, in Venice, or in London, or in Bruges, how many Venetian ducats, or English shillings, or Flemish groten, could you have bought with your 100 gold florins?

Historians of the Middle Ages all have encountered this kind of question. Currency fluctuations were even wilder then than they are today – with our weak dollar, strong yen, and political mark. Back in the spring of 1982, when I was writing the early chapters of a book on medieval saints, it took me the better part of two weeks simply to figure out how much work a craftsman such as a cloth dyer would have had to do in order to pay that fine, and that didn't even begin to answer the other aspects of the question.

What to do? It was puzzles such as this that led me, together with Professor Martha Howell, to found the Medieval and Early Modern Data Bank, which we refer to by the acronym MEMDB. Two circumstances then occurred to propel the MEMDB out of its tentative initial stages and into rapid development. The first of these circumstances was the generous donation to the project by Dr. Peter Spufford of Cambridge University of 13,256 medieval currency exchange rate quotations that he had compiled for his Handbook of Medieval Exchange. These exchange rates range in date from 1106 to 1500, and cover all of Europe,

[1] Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., U.S.A. on May 26-29, 1988

Byzantium, the Levant, and North Africa. They now form the pilot data set of the Bank, as I shall demonstrate in a few minutes.

The second fortunate circumstance for MEMDB was that in 1985 it attracted the attention of the Research Libraries Group (RLG). The RLG, which is based at Stanford, is a consortium of major research institutions. Through its on-line information system known as RLIN (Research Libraries Information Network), the RLG links libraries, museums, record offices, and research institutes in North America and Europe to provide access to, at present, a total of some 28 million bibliographic records. The RLG was interested in expanding its scope to include non-bibliogrpahic information, and offered to co-sponsor the MEMDB. As a co-sponsor, the RLG has provided three crucial services to the project: it assisted us in obtaining substantial funding for two years of full-time work; it developed the necessary computer software; and it offered to distribute the completed version of MEMDB on-line through its RLIN system.

Mention must also be made of Dr. Martha Carlin, who directs the day-to-day operations of the Medieval and Early Modern Data Bank, and who prepared the leaflet I trust you find before you. While I am here for a pleasant day in Washington, she is mired in the real work of adding new data sets to the Bank.

The MEMDB has now completed its first year of fully-funded work, and the results are exciting. We have succeeded in creating an up-and-running prototype version of the Bank that is designed to be used on personal computers, with Peter Spufford's 13,256 exchange rates as its data set. Copies of this prototype will be available publicly beginning this September, and I would like to spend the next few minutes demonstrating the system to you.

MEMDB was truly designed with novice computer users in mind. There are explanatory screens that provide all the basic information on how to scan indexes, find data entries, custom-design tables, and keep desired results. For example, the welcome screen directs new users to an introductory explanation screen, and reminds more experienced users of the most common commands (FINd, SCAn, and STOp), and of the two handy help screens that EXPlain ACTIONS and EXPlain INDEXES.

Let us assume that we want to find out if Spufford's collection of exchange rates contains any entries that refer to London between the years 1300 and 1450. If we were to look in the index to his book, we would see that there are 29 scattered pages that contain references to London, and we would have to examine each of these pages individually in order to see if they concern the years 1300-1450. To conduct this same search in MEMDB, we first scan the general index by typing

SCAn LONDON.

This quickly shows us that there are 199 entries in Spufford's data that refer to London. To see them, we simply

press the F6 key

After a brief wait, we may

[scroll down the entries to find those that occurred in 1300-1450.] 31.

Let us examine one of these entries in detail, for example, by displaying entry number

[DISplay] 31.

We notice that several of the fields have right-arrows; they indicate that the displays of these fields have been truncated. Similarly, there isn't enough room on a computer screen to display all the fields simultaneously. To see these off-screen fields, we

scroll to the right

until a message appears telling us that we have reached the rightmost edge of the table. In order to see a full-length display of all the fields in an entire line, we simply

press the F3 key

This will show not only all the fields of this entry, but also the entire contents of each field, without truncation. In entry number 31, for example, we see that on July 30 1387 a commercial exchange took place at London, in which 1 English pound sterling was exchanged for 26 shillings and 8 pence of Scotland. This amount is also converted for us into its decimal equivalent of 26.6667 shillings.

This full-length or "long" display also tells us that there is a background text to this entry, which we

press the F4 key to display.

As we

scroll down through the background text,

we see that it begins by describing the Scottish currency, and its relationship to English money; following which

[at the 9th screen down]

it contains a similar background history of medieval English money. If we

continue to scroll down, to the end of the text on English money

[7 more screens down],

we find ourselves at the beginning of Spufford's Introduction to his book, and if we were to continue on for about another 300 screens we would reach the Preface, at the end of which (about another 122 screens) we would finally reach the end of the background text.

In fact, each of Spufford's exchange rate entries has an appropriate background text, and each background text ultimately leads back into the Introduction and Preface. For example, if we

DISplay 30 BACK

to look just momentarily at the background text for entry number 30, which involves Florentine florins and English shillings, we see that it provides background information that is specific to these currencies. This feature is of fundamental importance. MEMDB provides not merely tabular results, but makes available the full scholarly apparatus behind each specific datum.

Having looked at the background text, we might wish to look at the sources of entry number 31. To do this, we

ESCape, scroll down one item, and press the F5 key.

In this case, as we can see, the source is an entry in the Calendar of the Close Rolls, as cited by Spufford on page 212 of his Handbook.

If we wanted to know what other exchange rates were culled by Spufford from the Calendar of the Close Rolls, using the book, this would mean looking through each of the 13,256 entries. However, to conduct this same search in MEMDB, we simply type

FINd TITle ?CLOSE ROLLS?

and in about 10 seconds we see that ten of Spufford's entries come from this source. Notice that instead of entering the entire title, I used question marks as "wild cards" to take the place of all other words and punctuation. This is convenient not only when searching for books or other types of entries with long names, but

also when one is uncertain about the spelling or wording of an entry.

To see what entries were taken by Spufford from this source, we simply

press the F6 key

to get a quick display. Obviously, what we have done here is to convert a text into a hierarchical data base, while retaining its textual features.

Perhaps at this point a user might wonder whether Spufford also consulted the Calendar of Patent Rolls or the Calendars of State Papers for his book. We can find out easily by using the handy "wild card" feature, the question mark, to type

FINd TITle CALENDAR?

This shows us that indeed he used

[scroll down, one by one:],

the Calendars of State Papers Milanese, State Papers Venetian, and Patent Rolls, as well as the Close Rolls.

Now, perhaps, the spirit of bibliographical inquiry is really moving us, and we want to see if Spufford has consulted certain other works, for example, the work of Alan Stahl, director of the American Numismatic Society. To do this we check the author index, by typing

FINd AUThor STAHL?

and in short order we find that there are 37 relevant citations. Again, we can quickly see these by

pressing the F6 key.

Thus far we have seen how one can search indexes of places, authors, and titles, create result sets of entries, and look at background texts and sources of individual entries. Now we can explore three more MEMDB facilities: reviewing and retrieving searches that we have already done, and custom-designing the tabular displays. To do this let us begin by typing

REView

This displays for us a listing of all the searches performed during this working session. We can retrieve an earlier search, such as the entries on London, by

putting the highlight bar on that search and pressing the F7 key.

Suppose that we wish to redesign the format of this set of results, so that only a few of its elements are displayed in full, and the remainder suppressed so that they do not clutter up the screen. To do this we use MEMDB's

TABle

routine to custom design our own table format. Let us say that we wish our table to display only the date, the place, and the type of exchange.

[TABle NEW DATe 22 PLAce 25 CATegory 25],

We type these field names in the sequence in which we wish them to appear on the screen, following each field name with the number of spaces that we wish to allot to it. Here, for example, we are assigning 22 spaces to the date field, 25 spaces to the place field, and 25 spaces to the category field. A sample line of our new table design is then displayed in the centre of the screen. If we are satisfied with the new design, and wish to use it to display result sets, we can keep it by typing

TABle KEEp

followed by an appropriate table name, such as

DATE-PLACE-CATegory

Now our result sets will be displayed in this new table design, until we change it again. For example, if we conduct a different search, for entries concerning the city of Paris,

[SCAn PLAce PARIS; F6],

we find that new searches, too, are displayed in this format. Up to 36 different table designs may be saved and retrieved for use at any time. To use a previously stored table design, we can first list all the available table names by typing

TABle REView,

and then select a table format by typing

TABle USE

followed by the name of the table we wish to use, for example,

STD,

if we wish to return to the standard table format used by MEMDB, or

TABle USE NO-DATE

(a table format that we designed and kept the other day) if we wish to use a table format that doesn't display the date.

As a final demonstration of the many wonders of MEMDB, let us return to my original 64,000-florin question: what could you exchange your 100 florins for in 1368? This would be a nightmarish search in the printed book: we would have to look at each of the 13,256 entries in order to find those that dated from 1368. In MEMDB, however, nothing could be simpler: we simply type

FINd DATe 1368

and then see

at entry number 1

that a Florentine florin could be exchanged at Toulouse for 12 gros tournois of France;

[move highlight bar down to entry no. 10]

or at Antwerp, for 27.0833 Flemish groten; or

[move highlight bar down to entry no. 11],

at Aachen, for 33 schillings of Cologne;

[move highlight bar down to entry no. 35]

or, in Rome, for 27 Flemish groten; or

[move highlight bar down to entry no. 37]

in Sicily for 6 Sicilian tari; or

[move highlight bar down to entry no. 38]

in England, for 3 shillings; and so on.

[STOp using MEMDB.]

That's the end of our on-screen demo; I'd like to conclude now by telling you something about the future of MEMDB. Essentially, MEMDB is a continually-expanding computer-based reference library for medieval and early modern historians. When MEMDB becomes an on-line facility, accessible through the RLIN system, we will be able to expand the Bank's scope to include virtually any scholarly compilation of data that can be presented in a tabular format. Thus we will include data on such subjects as wages and prices, household size, mortality, wealth, manufacturing, property-holding, and nutrition, to name only a few categories; drawn from such sources as taxation records, wills and inventories, parish records, import and export

records, household and estate accounts, prosopographical studies, archaeological reports, and so forth. Always, we will retain full textual background for individual data items.

We will also provide on-line reference aids, such as glossaries of weights and measures, gazetteers of Latin and vernacular place names, and calendars of dates. In addition, we will be a clearing-house for information about data bases that are in progress or are held by other institutions, both here and abroad. Finally, MEMDB will serve as a prompt and effective means of publication for scholars whose data bases are too costly to publish in print, and too clumsy to publish in microform. We will be able to incorporate such data bases into MEMDB's master data set, so that they will be simultaneously and mutually searchable, while at the same time each individual entry will retain its own original documentation, as we saw with Spufford's data. We will also gladly accept data collections offered to us that are valuable, but which do not particularly complement other material currently in the master data set. In such cases we will archive the data off-line, and will make these studies available to users through our Rutgers office.

I have tried to show you something about the Medieval and Early Modern Data Bank. Please pick up a leaflet and by all means do write or telephone us with any ideas and suggestions you would like to share with us. Thank you.□

Academic Libraries and Collection Development of Nonbibliographic Machine Readable Data Files

by Daniel C. Tsang[1]
Social Sciences Librarian/Bibliographer
University of California

It has been over two decades since Phil Converse (1964) and Ithiel de Sola Pool (1965) called on librarians to take the initiative in providing service for machine-readable data files (mrdfs) arguing that such materials logically belong in libraries. Since that clarion call, others have echoed their vision, as librarian Howard White (1974) discusses in his pathbreaking dissertation on social science datasets.

One of the pioneers that White documents, Ralph Bisco (1967, 17) in a speech delivered at the opening of the University of Florida's graduate research library, noted that two librarians (one of them Herbert Ahn, a colleague of mine, and then Systems Librarian at the University of California, Irvine) were members of a subcommittee of the newly established Council of Social Science Data Archives, a federation of data archives, that sought to improve access to social science datasets.

Another pioneer was Karl Pearson, whose 1968 University of California, Los Angeles library science thesis was described by White as "the fullest statement to date on libraries and numerical data sets" (White, 1974, 30-31). Yet another visionary was Jack Dennis, who envisioned the eventual assimilation of the data archive by the library after an "initial period of cooperation" (cited in White, 1974, 36; see also Adams and Dennis, 1970, 57-58).

Dennis, Linton Freeman and Robert Hayes were members of a Council on Social Science Data Archives committee that visited Northwestern University's Intersocietal Information Center in 1968, and subequently issued a report which recommended placing Northwestern's data archive in the university library as the "best place to have such a facility because of its central location, its interest in computer-oriented approaches, its knowledgeable personnel, its general policy of serving all people in the university, and the apparent availability of space

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., May 26-29, 1988

necessary to house such a facility" (cited in White, 1974, note 65, 213). But that did not happen.

According to another observer, "[s]ince libraries did not regard such data files as within their collection and service parameters, data archives, to a large extent were operated independently of libraries; libraries often neither managed these archives nor referred their clients to them... By 1984, more libraries were accepting responsibilities for the collection, preservation, and dissemination of nonbibliographic machine-readable data bases as a legitimate activity" (Hernon, 1986, 341), although most of these efforts dealt with online databases.

In the last twenty years or so, the library world has devoted several special issues of scholarly journals to the topic of nonbibliographic data files (White, 1977; Claydon and Soergel, 1982; Heim, 1982). In IASSIST itself, more and more members are traditional librarians beginning to provide mrdf service. At this conference, there is even a workshop on integrating mrdfs into traditional library service.

Traditional libraries struggling with what to do with mrdfs have in the past been rather conservative in their move to integrate mrdfs into the library. For instance, until recently, most have avoided cataloging such materials, except perhaps codebooks.

There has been an absence of any overall policy that would state clearly the role mrdfs play in library collections.

In part, that is perhaps due to the traditional library's mrdf 'phobia', even as more and more libraries are overcoming 'computer-phobia'. For example, many libraries are actively acquiring CD-ROM products, or subscribing to online services, and patrons are enthusiastic about searching Infotrac or BRS After-Dark. But mrdfs -- by which I mean data files used on mainframes for computerized statistical analysis -- are still considered for the most part a bothersome format.

In this paper, I want to focus on the collection development implications of having mrdfs in a traditional library setting. I shall limit my definition of mrdfs to nonbibliographic files accessible through mainframes, and not deal with the acquisition or collection of CD-ROMs and the like.

Traditional library literature is not much help in this area; in Library Literature (February 1988 issue), under "Collection development," the researcher is referred to "Libraries -- Book collections!" More diligent research will locate a few essays on or brief references to the topic, mostly on the process of how to locate data, i.e., data acquisition. Again, Bisco (1964), is a pioneer, writing twenty-four years ago about the "complex" process involved in acquiring new data. Robert Mitchell (1964, 90-91) also is an early essayist on the acquisition of Third World datasets. David Nasatir (1977), in a brilliant essay, expounds on the joys and tribulations of "Stalking the Wild Data Set" at home and abroad. He also provides the most extensive discussion on the "Data Acquisition Function," in a UNESCO report entitled Data Archives for the Social Sciences (1973).

In his Ph.D thesis White (1974), focuses on an analysis of purchase orders at a number of social science data suppliers, and concludes with a call for libraries to actively purchase codebooks, but not datasets.

Betty Yantis (1980) and John Nixon (1980) both write about the acquisition of data at the University of Nevada's Center for Business and Economics Research. Robbin (1977), writes about the "pre-acquisition process." Ray Jones (1982) describes the variety of governmental and academic datasets acquired at the University of Florida Libraries, and Pope (1984) details the committee process involved in the acquisition of new datasets at that institution. Ann Gerken

(1984, 9), then at the Cornell Institute for Social and Economic Research, notes the variety of sources used for data collection, and adds that "[d]etailed collection development policies are developed in collaboration with faculty, librarians, and members of the Institute." Then U.S. Archivist Bob Warner and Francis Blouin (1980) as well as Charles Dollar (1980) address the complex issue of appraising mrdfs. Chiang (1986, 68) reports that Cornell Library collects "both microcomputer and mainframe accessible data."

A data archive's implicit collection development guidelines may evolve into something more explicit. Laine Ruus (1982a, 399-400), then at the University of British Columbia Data Library (created as a joint library-computing facility venture), concedes that as is commonly the case, "the collections policy is rather vague and ad hoc. The original mandate made only one stipulation regarding collections — that the Data Library 'develop collections...in accordance with the academic requirements of the University, in parallel with the policies of the Computing Centre and the Library'." Nonetheless, a "policy has evolved over the years" which can be summarized as follows: "the Data library will collect automatically all significant Canadian data files such as census data, election studies, and other major social surveys... All other MRDF are acquired on request, tempered by considered need, potential for furture use, and of course, budgetary constraints. In addition, the library will function as a data archives in the sense that an attempt is made to acquire any original MRDF produced by local researchers, or offered for deposit by outside researchers (depository MRDF) and every effort is made to ensure that these are maintained for prosterity."

Ruus (1982a, 400) has not found it necessary to limit datasets to particular disciplines, but she will not acquire datasets that cannot be made available to all academic users, or where individual privacy is violated. Nor will she maintain mrdfs that "lack adequate documentation or are so 'dirty' as to be useless for secondary analysis." UBC's Data Library's collection development policy is well delineated in the "Data Library Procedures Manual," which offers a section on the "Care and Feeding of the MRDF Collection," and includes within that an acquisitions policy (Ruus, 1982b; Henderson, 1988). Its policy may well be the most explicit of all such collections, even to the extent of considering previous use patterns as shown by tape mount statistics.

A 1984 survey by the Association of Research Libraries found that of 34 responding libraries, only four had drafted a collection development policy statement for machine-readable data bases (Association of Research Libraries, 1984, 3). The four are not further identified.

From my own informal, admittedly small sampling of a handful of data archives, it appears that most do not have written collection development policies. At Simon Fraser University, Walter Piovesan (1988) has "found no real need to actually formalize a policy... Not having a fixed budget makes it hard to develop a collection policy. If you have a budget, then you will need to 'prioritize." Ann Janda (1988) at Northwestern University data archive, which is operated by the Academic Computing and Network Services rather than the library, also does not have a "formal collection development policy for MRDF... at least not yet," with orders "need and demand driven." According to Janda, "This has worked fine as we have been working largely in the realm of ICPSR data requests." For users requesting non-ICPSR data, she does invoke certain principles: The data requested must be of a "fairly general nature" and is potentially going to be used by more than one project or user; and the user should share the cost of the purchase." At Yale, JoAnn Dionne (1988) reports that the Social Science Data Archive "doesn't have a formal collection policy for mrdfs." She generally "buys only when a user

requests a data set and only then only within certain dollar amounts -- but that depends on anticipated use." She will not spend more than $200 unless more than one person will use the data.

A 1983 report by a task force on nonbibliographic databases at the State University of New York, Albany calls for drafting a collection development policy that at first "should be limited to acquiring data sets on demand. The collection should also be limited to locally developed and/or locally used data" (State University of New York, 1983, 22).

In the University of California library system, work has begun on drafting local mrdf collection development policy statements. A draft policy, "Collection Policy Governing Machine Readable Data Files," dated November 11, 1986, for the Berkeley library, applies to all mrdfs, including software. In its present version, which may not survive in final form, it recommends substituting mrdfs for printed information only "with extreme caution," given the volatility of the information industry, the limited number of simultaneous users, and the need for staff assistance. It also cautions against acquiring mrdfs purely as a depository function, and urges that all collection be evaluated against other potential acquisitions and weighed against other uses of book monies.

At University of California, Davis, library staff have been discussing recommendations of a committee on numeric and textual databases; a proposal that "collecting responsibility for all formats belongs to all selectors" received "wide support," as did a recommendation that the position of a coordinator for machine-readable resources be created. At the present time, "programming expertise sufficient to access large datafiles on mainframe computers is not required, although the candidate may need to develop this ability if the need arises" (University of California, Davis, Library, 1987). At University of California, San Diego, Jim Jacobs (1986), also has drafted a statement on collection development, describing the collection there as "being built passively in response to requests for machine readable data from faculty. There is no active program for acquiring data in anticipation of possible future needs."

The Research Libraries Group, comprising the nation's top research collections, has also begun working on a mrdf collection management study (Jones, 1988).

I suspect a major reason for the flurry of activity among collection development librarians is the proliferation of CD-ROM products. A policy is needed before libraries become inundated with new technological products. I suspect mrdfs on tape are the least of most librarians' worries.

Nonetheless, before proceeding, it may be useful to pause and reflect on why a collection development policy would be productive. After all, most data archives appear to have survived without such written policies!

Several reasons come to mind. First, more and more libraries are being run like corporations, and hence, wanting a written policy for every procedure may just be the application of good management pracatices.

Also, we must realize that our individual tenure as data archivists or data librarians -- whatever we may hope -- is finite. At a recent meeting on mrdfs with the technical services librarians in my library, the Head of Acquisitions turned to me and said plaintively, "Dan, you could get run over by a car tomorrow!" In other words, what happens when I am gone? Will my replacement be able to figure out what was done? To be sure, through our daily work, every good librarian becomes a repository of obscure facts and important information. That is unalterable. Not everything can be written down, or passed on easily to the next generation! But a collection development policy

would be one way in which to clarify on paper what past practice has been, and what future practice should be.

Another major reason is so that the collection can develop in an orderly manner, and not be subject entirely to the whims of a particular bibliographer or researcher.

A well-written policy on collecting mrdfs might also protect the collection from any arbitrary change; at least, one could point to the policy to try to forestall any attempt to get rid of the collection!

In addition, such a written policy would be useful in training or orientating new staff in the library, as well as an aid in publicizing or explaining the collection, to users and potential donors.

At Irvine, the librarians who do the collecting are called bibliographers; at other libraries, they could be called "selectors." One immediate question we are confronting at Irvine as we merge mrdfs into the collection, is whose responsibility it is to select mrdfs? In theory, our bibliographers are responsible for all formats of materials. On paper this looks good; but in practice, this has primarily and almost exclusively meant traditional library formats such as books, periodicals, microform, video or audio tapes or films. As the Official Representative for ICPSR and the Social Sciences Bibliographer, I have become the *de facto* mrdf selector. The working solution we have implemented is that all mrdf selections will be passed through me just to see they are compatible with the hardware (or software) researchers use. But thus far, no one else has placed any orders.

To let all bibliographers select mrdfs may sound like heresy to a data archivist; but I would argue that if we are to integrate mrdfs into traditional library services, we must avoid stereotyping mrdfs as some weird format, and thereby perpetuating the segregation by format.

If the problem is lack of awareness or familiarity with mrdfs, then that surely can be remedied. Just as reference librarians are retooling for database searching, I believe that bibliographers can be educated about mrdfs. Instead of seeing this as an attack on one's turf, one might rather see this as an opportunity for others to contribute their expertise. For bibliographers are subject specialists who are responsible for working with faculty, and thus should be aware of the research needs on the campus within a particular discipline. There will soon be no way that one person can know or attempt to meet all the mrdf needs of all the disciplines on a campus.

In traditional libraries, collection policies for books and serial titles at major academic libraries generally are divided by subject and within each subject, by level of collecting. For example, at the University of California, Irvine (1983, 25-26), the levels are comprehensive ("all significant works" within a defined field), research (supporting doctoral and post-doctoral work), advanced study or beginning research (graduate and advanced undergraduate work), teaching or initial study (undergraduate curriculum), and the lowest level, basic information (non-curriculum-related).

Sections on mrdfs, then, could well be included within the individual subject chapters of a collection development manual, where mrdfs are an important part of a research or instruction program on campus. That, I believe, would be a long-term goal as mrdfs become futher integrated into traditional library services.

But it still would be helpful for the library to have an overall collection development policy on mrdfs, if only because of the processing and service implications any acquisitions entail.

Having a written collection policy does not mean it is engraved in stone. A policy must be

flexible and open to revision (Robbins, 1977, 25). No one policy can be written for all data collections. Local conditions will dictate what is important for that collection (Bernard and Jones, 1984, 98).

With that in mind, I would just like to focus on some important elements I believe such a policy should contain. Many of the ideas or categories are taken from a report on "Textual and Numeric Data Files" written by an ad hoc committee of the Librarians Association of the University of California (1983, espcially 13–14).

I have also garnered some ideas from an amazing book of abstracts compiled by the staff at the Correlates of War Project at the University of Michigan. Beyond Conjecture in International Politics (Jones and Singer, 1972), is a collection of abtracts of data-based research, systematically analyzed by a set of categories that are useful as we develop a collection development policy. Finally, others come from essays already cited.

Some important elements of a mrdf collection development policy for an academic library are:

1. Selection responsiblity. Who has the responsibility; all bibliographers? Or just the mrdf bibliographer?

2. Budget source: Who pays?

3. Level of collection activity (see above).

4. Subject Scope: Is the subject of relevance to research and instruction at the university?

5. Temporal domain: Is the time period covered of relevance to research and instruction at the university?

6. Spatial domain: Does the mrdf cover a region or location that is of relevance to research and instruction at the university? For example, Is the data collection a regional depository?

7. User need: Does the user need to manipulate data or just use manipulated data?

8. Uniqueness of data: Are the data available in print format? Is it necessary to get it in mrdf format? Are they only available in mrdf format?

9. Currency of data: Are the data from an ongoing study that will be quickly superseded by more recent revisions? Is it important to acquire quarterly updates or just annual cumulations?

10. Confidentiality of data: Is there a need to restrict personal or proprietary information in the dataset? Will acquisition violate privacy?

11. Physical format: Is the medium compatible with available hardware?

12. Software compatibility: Are the data accessible by software currently available? Or are they software dependent?

13. Documentation: Are the data supported by adequate documentation?

14. Data quality: Are the data sufficiently "cleaned" so that the data set can be added to the collection without further processing?

15. Access: Is the data set accessible to all users? Are there any restrictions? Is it accessible online?

16. Producer reliability: Is the distributor/producer of the data reliable? Are its products well regarded?

17. Historical importance: Is the data set worth preserving even if use is limited in the foreseeable future?

18. Ownership: Who retains ownership of the data set?

19. Levels of analysis: Does the data desired level of analysis?

For guidelines to the evaluation of a scientific data set (as opposed to a social science dataset), see Bruce Ewbank's "Comparison Guide to Selection of Databases and Database Services" (1982).

Drafting of a sound, collection development policy is a prerequisite before a library engages in a full-scale effort to acquire data files. Otherwise, bibliographers may well be forfeiting responsibility for selection to database service suppliers and vendors, or to the user community. Collections that are based entirely on demand without any clear policy may be uneven, lack depth or focus, and become unmanageable. As Bisco (1970, 282) pointed out almost twenty years ago, "there are notable gaps in the collections of archives."

At the very least, the process involved in drafting a policy can be used to bring together all bibliographers and other librarians so that it serves an educational and unifying function, and further help integrate mrdfs into a traditional library setting.

The challenge for those of us who straddle both the library and the archival worlds is to develop a collection policy that is not overly restrictive, but flexible enough to permit us as selectors the necessary leeway to develop our collections.□

References

Adams, Margaret O'Neill and Jack Dennis. 1970. Creating local social science data archives. Social Sciences Information, 9/2 (April), 51-60.

Association of Research Libraries. 1984. Nonbibliographic macine-readable data bases in ARL Libraries (Washington, D.C.: ARL, Office of Management Studies, Systems and Procedures Exchange Center). SPEC Kit 105.

Bernard, H. Russell and Ray Jones. 1984. The use of MRDFs in the social sciences: an anthropologist and a librarian look at the issues. In Association for Research Libraries, Nonbibliographic machine-readable data bases..., 94-99.

Bisco, Ralph L. 1967. The research library and data archives for social research. Speech prepared for the dedication of the Graduate Research Library, University of Florida.

————————. 1970. Summing up. In Ralph L. Bisco, editor, Data bases, computers, and the social sciences (New York: Wiley), 273-285.

Chiang, Katherine S. 1986. Computer accessible material in the academic library: avoiding the kludge. In Danuta A. Nitecki, editor, Energies for transition: proceedings of the Fourth National Conference of the Association of College and Research Libraries, Baltimore, Maryland, April 9-12, 1986 (Chicago: The Association), 67-69.

Claydon, Charles R. and Dagobert Soergel. Numeric databases. Drexel library quarterly, 18/3-4 (Summer-Fall).

Converse, Philip E. 1964. A network of data archives for the behavioral sciences. Public opinion quarterly, 28/2 (Summer), 273-286.

Dennis, Jack. 1977. The relation of social science data archives to libraries and wider information networks. In Howard D. White, editor, Reader in machine-readable social data (Englewood, Colo.: Information

Handling Services), 165–174.
Dionne, JoAnn. 1988. Electronic mail message to Dan Tsang, 8 March.
Dollar, Charles M. 1980. Machine-readable records of the federal government and the National Archives. In Carolyn L. Geda et al, editors, Archivists and machine-readable records, 79–88.
Ewbank, W. Bruce. 1982. Comparison guide to selection of databases and database services. Drexel library quarterly, 18/3–4 (Summer–Fall), 189–204.
Fischer, Jayne D., editor. 1980. Readings in business and economic research management: execution and enterprise. Vol. 1. Madison: University of Wisconsin, School of Business, Office of Research Administration.
Geda, Carolyn L. et al, editors. 1980. Archivists and machine-readable records: proceedings of the Conference on Archival Management of Machine-Readable Records, February 7–10, 1979, Ann Arbor, Michigan (Chicago: The Society of American Archivists).
Gerken, Ann E. 1984. Organizing a data library. IASSIST quarterly, 8/4 (Winter), 7–10.
Heim, Kathleen M., issue editor. 1982. Data libraries for the social sciences. Library trends, 30/3 (Winter).
Henderson, Jim. 1988. Electronic mail message to Dan Tsang, 18 May.
Hernon, Peter. 1986. Numeric data bases. Encyclopedia of library and information science, 40/suppl. 5, 339–354.
Jacobs, Jim. 1986. Collection profile. University of California, San Diego.
Janda, Ann. 1988. Electronic mail message to Dan Tsang, 11 March.
Jones, Ray. 1982. The Data Library in the University of Florida Libraries. Library trends, 30/3 (Winter), 383– 396.
———. 1988. Telephone interview, 6 May.
Jones, Susan D. and J. David Singer. 1972. Beyond conjecture in international politics: abstracts of data-based research (Itasca, Illinois: F.E. Peacock).
Librarians Association of the University of California. 1983. Textual and numeric data files: primary information sources in machine-readable form. Reprinted in Assocation for Research Libraries (1984), Nonbibliographic Machine-Readable Data Bases..., 5–18.
Mitchell, Robert Edward. 1965. A social science data archive for Asia, Africa and Latin America. Social sciences information, 4/3 (September), 85–103.
Nasatir, David. 1973. Data archives for the social sciences: purposes, operations and problems (Paris: UNESCO).
———. 1977. Stalking the wild data set: the acquisition of machine-readable social science data at home and abroad. Drexel library quarterly, 13/1 (January), 43–47.
Piovesan, Walter. 1988. Electronic mail message to Dan Tsang, 5 April.
Pool, Ithiel de Sola. 1965. Data archives and libraries. In Carl F.J. Overhage and R. Joyce Harman, editors, INTREX: Report of a planning conference on information transfer experiments, September 3, 1965. (Cambridge: MIT Press), 175–181.
Pope, Nolan F. 1984. Providing machine-readable numeric information in the University of Florida Libraries: a case study. In Ching-Chih Chen and Peter Hernon, editors, Numeric databases (Norwood, New Jersey: Ablex), 263–282.
Robbin, Alice. 1972. The pre-acquisition process: a strategy for locating and acquiring machine-readable data. Drexel library quarterly 13/1 (January), 21–42.
Ruus, Laine G.M. 1982a. The University of British Columbia Data Library: an overview. Library trends, 397–406.
———. 1982b. Data Library Procedures Manual. Excerpted in Association of Research Libraries (1984), Nonbibliographic machine-readable data bases..., 62–97).
State University of New York, Albany. 1983. Final report. Task Force on

Non-bibliographic Databases. Reprinted in Association for Research Libraries (1984), Nonbibliographic machine-readable data bases..., 22-56.

University of California, Davis, Library. 1987. Report of the Committee on Numeric and Textual Databases: summary of comments and recommendations for revision.

University of California, Irvine. 1983. Collection development policy for the University Library.

Warner, Robert M. and Francis X. Blouin, Jr. 1980. Some implications of records in machine-readable form for traditional archival practice. In Carolyn L. Geda et al, editors, Archivists and Machine-Readable Records, 242-248.

White, Howard Dalby. 1974. Social science data sets: a study for librarians. Dissertation (Ph.D, Librarianship). University of California, Berkeley.

————————, editor. 1977. Machine-readable social science data. Drexel library quarterly, 13/1 (January).

Yantis, Betty. 1980. Data acquisition, storage and management in a new research center. In Jayne D. Fischer, editor, Readings in Business and Research Management, 117-119.

UNIVERSITY OF ALBERTA

1989 SUMMER INSTITUTE

A Workshop on Managing 1986 Computer-readable Census Files

Instructors: Laine Ruus, University of Toronto Library
Walter Piovesan, W.A.C. Bennet Library, Simon Fraser U.
Chuck Humphrey, Data Library Coordinator, U of Alberta

Sponsor: The Population Research Laboratory and University of Alberta Computing Systems in conjunction with the Summer Institute

Date: Saturday, June 24th, 1989
(during the Canadian Library Association annual conference)

Time: 9:30 a.m. – 4:00 p.m.

Place: University of Alberta

Cost: $25

Computer-readable files of 1986 Census of Canada data at very small geographic levels as well as individual level data, are being purchased by universities and colleges across Canada, through a Consortium led by CARL/ABRC (Canadian Association of Research Libraries). These files provide more flexible access to census information than the standard printed publications. Few academic libraries currently have the systems or experienced staff to handle computerized data, especially of this volume and complexity.

The workshop will introduce basic management techniques needed to work with these computer products. The topics to be covered will include: content and structure of the 1986 computer-readable census data, technical processing of the data within existing library/computing facilities, cataloguing standards for computer files, integration of census data into traditional reference services and materials. Examples of common uses of the data, limitations of the data, use with popular statistical packages, the organization of the files on computer tape and accompanying documentation, and data retrieval techniques will also be discussed.

For further information, contact Ilze Hobin, Population Research Laboratory, University of Alberta, phone (403) 492–4659 or e-mail: USERINS7@UALTAMTS.

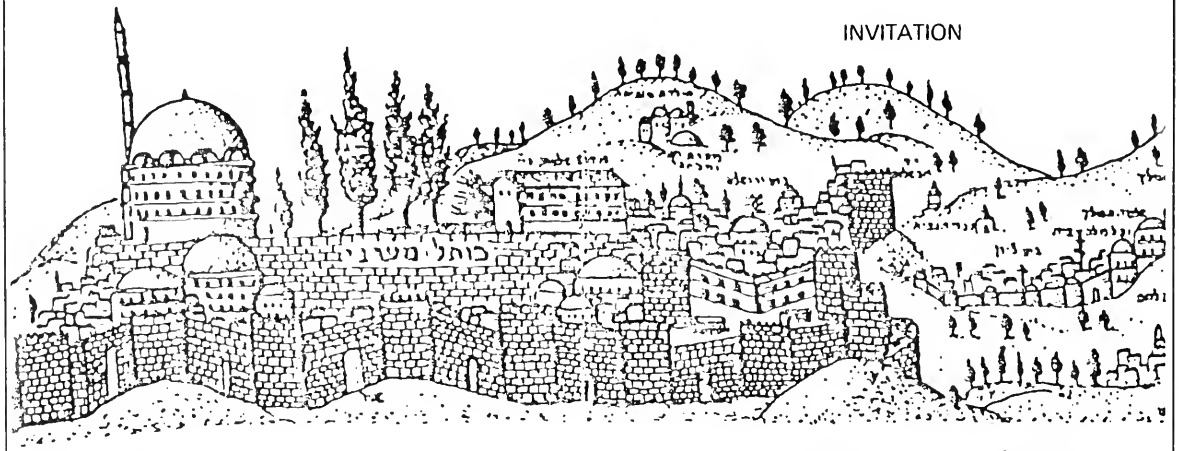
האוניברסיטה העברית בירושלים
הפקולטה למדעי החברה
מאגר הנתונים

IFDO / IASSIST 89
JERUSALEM, MAY 15-18, 1989

ערכם של נתוני מחקר למגזר הציבורי ולעסקים

VALUE OF RESEARCH DATA FOR GOVERNMENT AND BUSINESS

INVITATION



כנס בינלאומי של מאגרי נתונים
הר הצופים, ירושלים, 18-15 במאי 1989

IFDO — International Federation of Data Organizations
IASSIST — International Association of Social Science Information Service and Technology

LISA

Library & Information Science Abstracts

- **international scope and unrivalled coverage**
  LISA provides English-language abstracts of material in over thirty languages. Its serial coverage is unrivalled; 550 titles from 60 countries are regularly included and new titles are frequently added

- **rapidly expanding service which keeps pace with developments**
  LISA is now available monthly to provide a faster-breaking service which keeps the user informed of the rapid changes in this field

- **extensive range of non-serial works**
  including British Library Research and Development Department reports, conference proceedings and monographs

- **wide subject span**
  from special collections and union catalogues to word processing and videotex, publishing and reprography

- **full name and subject indexes provided in each issue**
  abstracts are chain-indexed to facilitate highly specific subject searches

- **available in magnetic tape, conventional hard-copy format, online (Dialog file 61) and now on CD-ROM**
  Twelve monthly issues and annual index

Subscription: UK £157.00
Overseas (excluding N. America) £188.00
N. America US$357.00

Write for a free specimen copy to

Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

CURRENT RESEARCH
in library & information science

- **CURRENT RESEARCH** is an international quarterly journal offering a unique current awareness service on research and development work in library and information science, archives, documentation and the information aspects of other fields

- **The journal provides** information about a wide range of projects, from expert systems to local user surveys. FLA and doctoral theses, post-doctoral and research-staff work are included

- **Each entry provides** a complete overview of the project, the personnel involved, duration, funding, references, a brief description and a contact name. Full name and subject indexes are included

- **Other features include** a list of student theses and dissertations and a list of funding bodies. Each quarter, an area of research is highlighted in a short article

CURRENT RESEARCH is available on magnetic tape, as well as hard copy, and can be searched online on File 61 (SF=CR) of DIALOG

Subscription: UK £86.00
Overseas (excluding N. America) £103.00
N. America US$195.00

Write for a free specimen copy to

Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

The International Association for Social Science Information Services and Technology (**IASSIST**) is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and /or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST.**

Membership fees are:
Regular Membership. $20.00 per calendar year.
Student Membership: $10.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription: $35.00 per calendar year (includes one volume of the Quarterly)

Membership form

I would like to become a member of IASSIST. Please see my choice below:

- $20 Regular Membership
- $10 Student Membership
- $35 Institutional Membership

My primary Interests are:

- Archive Services/Administration
- Data Processing/Data Management
- Research Applications
- Other (specify) _____

Please make checks payable to IASSIST and Mail to :

Ms Jackie McGee
Treasurer, IASSIST
% Rand Corporation
1700 Main Street
Santa Monica

Name/phone

Institutional Affiliation

Mailing Address

City

Country/zip/postal code

SANTA MONICA
MAY 1 '89
CALIF.
0.90
METER
8201104

SPECIAL FOURTH-CLASS
BOOK

LIBRARY

MAY 15 1989

School of Library Science
University of North Carolina